Essays in Applied Microeconomics

with a Focus on Vocational Education and Training

Inauguraldissertation zur Erlangung der Würde eines Doctor rerum oeconomicarum der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität Bern

vorgelegt von Manuel Aepli

von Bern

2020

Die Fakultät hat diese Arbeit am 20.02.2020 auf Antrag der beiden Gutachter Prof. Dr. Gerfin und Prof. Dr. Mühlemann als Dissertation angenommen, ohne damit zu den darin ausgesprochenen Auffassungen Stellung nehmen zu wollen.

To my parents and all my other teachers

Preface

This thesis consists of three chapters, each derived from an individual paper. Although each of these chapters deals with a unique research question, there are similarities that are strong between Chapters 1 and 2 and weak regarding the third. First, all chapters raise questions in the field of microeconomics and touch on the Swiss system of Vocational Education and Training (VET). Second, all chapters apply quantitative methods widely used in the field. And finally, all chapters are embedded in the Swiss context. Whereas this *Swissness* merely concerns the data in chapter three, it additionally includes the outcome of interest and the fundamentals of the empirical methodology in chapter one and two.

Both Chapters 1 and 2 investigate firms' engagement in the Swiss VET system. Within this system firms' voluntary participation is crucial – they help design curricula, hire apprentices, pay their wages, and are responsible for most of their training - and remarkable: Setting a world record, Switzerland's dual VET system accepts around 60% of all pupils after compulsory schooling each year. It thus seems fair to say that Swiss firms bear a large proportion of the investments needed to secure their own future skill demand.

Chapter 1, which is joint work with Andreas Kuhn, investigates what happens to this voluntary engagement in the skill formation process if firms are permitted to secure their skill demand from another source: we focus on immigration. In recent years, this channel has changed substantially, both quantitatively and qualitatively. Through mutual agreements, foremost the Agreement on the Free Movement of Persons implemented in 2002, the labor markets of Switzerland and the European Union have increasingly integrated, and the non-Swiss workforce has grown by roughly 50% or 600,000 workers since 1995. Meanwhile, most immigrant workers today hold a vocational or tertiary degree, whereas in the 1990s fewer reached educational attainments higher than compulsory schooling. Overall, we hypothesize that this enlargement of skills provided by the non-native to the Swiss labor market may incentivizes firms to substitute their investments in VET by hiring immigrant workers.

To examine this claim empirically, Chapter 1 focuses on crossborder workers, who work in but live outside Switzerland. Today, crossborder workers account for 6% of the total Swiss workforce and 20% of the immigrant workforce. Moreover, their numbers doubled between 1995 and 2018, an increase twice as large in relative terms as the simultaneous increase in resident immigrant workers. To understand Switzerland's particular attraction for crossborder workers, it is first worth picturing a map of Switzerland's distinct language regions. The triangle on the left pointing towards France forms the French-speaking part. The much smaller triangle hanging upside down into Italy covers the Italian-speaking, and the rest, mostly bordering Germany and Austria, the German-speaking part. One notices that all three main Swiss language regions border countries in which they are the sole official language. Evidently, language is not a major barrier at the Swiss border within the institutionally relatively well integrated European labor market, which together with the comparatively high Swiss wages makes Switzerland attractive for workers from neighboring countries. Second, Switzerland's comparatively high cost of living together with assumed personal preferences for residing in one's country of origin makes it relatively unattractive to work *and* live in Switzerland.

Quite obviously, the resulting high numbers of crossborder workers are not evenly distributed across Switzerland. Swiss firms' opportunities to employ them is substantially constrained by their distance from the border. The empirical approach presented in Chapter 1 exploits this setting by comparing firms close to the border with large and firms far from the border with limited access to crossborder workers that are otherwise similar. Overall, we find that the increase observed in crossborder workers between 1995 and 2008 led to a decrease of about 3,500 apprenticeship positions, corresponding to roughly 2% of the total number of apprentice positions. Although the exact channel through which this substitution works remains ambiguous, policy makers designing institutions in either immigration or VET might want to pay attention to this trade-off, especially because it involves two goods that are in general positively valued by many employers.

Chapter 2, which is joint work with Andreas Kuhn and Jürg Schweri, sticks with firms' investments in VET and, moreover, also investigates the spatial distribution of it across Switzerland. The focus lies on the varying proportions of firms providing apprenticeship positions across Switzerland's language regions (that you remember when recalling the map pictured above). Note first that we do not claim there is any direct link between different languages and different levels of firm engagement in VET. However, the distinct languages in Switzerland also maintain cultural differences within a small country, despite the fact that national institutions are well accepted and people increasingly mobile. It may be due to exclusive communication with same-language peers, selective media consumption, or varying exposure to Switzerland's neighboring countries with whom each Swiss language region forms a distinct supranational linguistic region: French speakers are most open to immigration and international cooperation and eat more meat than their German-speaking counterparts. German speakers have a more traditional understanding of gender roles than their French-speaking counterparts and donate the most for charity of all linguistic groups. Italian speakers use public transport the least and value leisure more than their German-speaking counterparts.

In Chapter 2, we focus on a clear discontinuity that analysis of voting results reveals among the language-cultural regions constituting Switzerland: Whereas French and Italian speakers approve of strong state involvement, for example in the health insurance sector, for pension schemes, and in the VET system, German speakers prefer private engagement over the states in the same domains. From this starting point, we ask whether a favorable attitude towards private engagement expressed at the ballot box is actually accompanied by higher levels of the privately provided good apprenticeship. Chapter 2 reveals the answer to be yes: firms located in German-speaking municipalities are about 10% more likely to train apprentices than firms in very nearby French- and Italian-speaking municipalities. Altogether, we argue that norm-guided behavior is a complementary explanation for why some firms train apprentices and others do not. One can draw two policy implications from this finding. First, persistent norms might strengthen the sustainability of the Swiss VET system against potential shocks to firms' cost-benefit ratio. Second, behavior bound by norms might hinder the export of a Swiss-style VET system with its strong focus on firm engagement to other countries even if they set up the institutional framework to foster it.

Chapter 3 focuses on the adult labor market, where skills acquired, e.g. in the VET system, are applied. Many economists claim that recent rapid technological change penetrating the labor market has shifted firms' skill demand and altered the nature of jobs. Given individuals' skills, these demand shifts potentially foster horizontal skill mismatches, such that someone's acquired skills do not match the skills needed in their current occupation. In line with previous mismatch literature, Chapter 3 shows that such horizontal skill mismatches are a multi-faceted phenomenon. First, whereas only about half of all individuals work in exactly the occupations that they learned formally, the degree of mismatch among the other half varies widely. Second, many individuals actually realize wage gains when becoming mismatched; this suggests that objectively identified mismatches are not bad per se. Based on these general findings and on the task-based approach, I hypothesize that horizontal skill mismatches are harmful to the wages of individuals who mostly hold skills substitutable by new technology, whereas they are not harmful in general. I account for this heterogeneity of mismatches in the empirical analysis of Chapter 3 by exploiting detailed occupational task data to measure the strength of mismatches and to focus on mismatches presumably caused by skill demand shifts due to new technology. The main result yields a wage penalty of roughly 12% for mismatched individuals with high shares of substitutable skills. Applying other methods to the same dataset suggests that objectively identified horizontal mismatches have zero wage implications on average, even after accounting for unobservable individual characteristics. From a policy perspective, it thus seems important to bear the heterogeneity of the mismatch phenomena in mind; otherwise, revealed average effects might mask negative effects on certain subgroups. In this spirit, I estimate mismatch wage penalties for different educational subgroups including VET diploma holders.

And in this sense, all chapters of this thesis deal with the Swiss system of vocational education and training, a system that is recognized by many as one of the key contributors to the country's economic success. However, as this thesis shows, it is also a system that contains frictions, even contradictions at first sight, a system regularly challenged by pupils entering it and labor markets demanding its outcomes, and therefore a system that must remain agile. Thus, the recognition that the Swiss VET system receives from inside and outside the country must motivate constant reflection, adjustment, and amelioration, and should never tempt to rest on its laurels. I hope this thesis plays its modest role in fulfilling this purpose.

Bern, January 2020

Acknowledgement

First of all, I thank Michael Gerfin, the supervisor of this thesis, for his helpful comments and guidance during the last three years. Special thanks go to Jürg Schweri, co-author of the second chapter and my supervisor at the Swiss Federal Institute for Vocational Education and Training, for his support during the writing process, for sharing his immense knowledge on the Swiss educational system with me, and for providing me with enough time to complete my thesis. Moreover, I thank him for giving me the freedom to pursue numerous paths on my own, many of which were dead ends but some of which led to this thesis. I thank Andreas Kuhn, co-author of the first and second chapters, a lot for countless fruitful discussions at the office or during coffee breaks, for always helping me with methodological issues, and for forcing me to dig deeper again and again. I am grateful to the Swiss Federal Institute for Vocational Education and Training for giving me both the chance to write this thesis and the opportunity to work on many other interesting projects at the same time. This thesis also benefited from six doctoral seminars of the Swiss Leading House for Economics of Education at the University of Zurich. All seminars delivered interesting insights into the latest state of research and taught me valuable methodological tools; I therefore thank the participants, the organizers, and the visiting lecturers. Finally, I am thankful to Sally Gschwend and Simon Milligan for proofreading the present thesis.

Contents

List of Figures

List of Tables

Chapter 1

Open Labor Markets and Firms' Substitution between Training Apprentices and Hiring Workers

joint work with **Andreas Kuhn**

1.1 Introduction

One of the key objectives of the European Union has been the integration of its member countries' labor markets. The main rationales of these efforts is to spur gains in innovation and productivity and, consequently, growth (e.g. Hunt and Gauthier-Loiselle, 2010; Peri, 2012). However, economic theory also predicts distributional effects from immigration, at least in the short run. This implies that some groups will presumably suffer from immigration, even when its overall effect on the host country is positive (e.g. Bansak *et al.*, 2015; Borjas, 2014). Consistent with this, several empirical studies have documented that recent waves of immigration have been accompanied by a rise in right-wing parties that seek to restict further immigration in various European countries (e.g. Brunner and Kuhn, 2018; Edo *et al.*, 2019; Halla *et al.*, 2017).

In this chapter, we estimate, against this broader background, employers' short-run substitution between the training of resident apprentices and the hiring of cross-border workers (hereafter CBWs), i.e. nonresident immigrant workers who live in one of the neighboring countries but who work regularly in Switzerland. Prospective apprentices are not yet in the labor market, they are not organized or politically represented, and they therefore have low bargaining power. They may thus be especially at risk of being substituted with CBWs. Moreover, training apprentices is costly for employers: average training costs over the full training period equal almost 100,000 Swiss francs, equivalent to about 1.28 times the annual median wage in 2016. Consequently, we hypothesize that access to CBWs will decrease the costs of hiring skilled workers and will therefore lead some firms to substitute away from training their own workforce to hiring more CBWs. Similar to Kerr *et al.* (2015) and others, we focus on employers' decisions using comprehensive firm-level data from the Swiss business census. We believe that the substitution of training with hiring workers by employers is a relevant topic from an educational-policy perspective for two reasons. First, apprentices are quantitatively important within the Swiss educational system: around two thirds of young people attend some kind of apprenticeship after compulsory schooling; and second, the Swiss apprenticeship system is often viewed as one of the key sources of the country's economic success.

We focus on substitution with CBWs for both substantive and methodological reasons. First, and in contrast to many immigrants both working and living in Switzerland (resident immigrant workers, henceforth[1]), CBWs presumably have skills similar to those of natives, and they have the advantage that they usually speak the same language as natives (e.g. Stöhr, 2015). We therefore expect that substitutability between apprentices and CBWs will be higher than between apprentices and non-native workers in general, many of whom are low skilled and/or do not speak any of the country's official languages.[2] Second, CBWs are quantitatively relevant

[1]This group also includes persons born in Switzerland with non-native parents, because citizenship is not automatically granted for persons born in the country.

[2]In general, immigrant inflows may have very different effects, depending on whether they are substitutes or complements with natives (e.g. Peri and Sparber, 2009). For

in the Swiss labor market: CBWs made up about 4.8% of the total workforce in 2008 (i.e. the year which marks the endpoint of our empirical analysis), representing about 18.2% of the total foreign workforce in that year. Moreover, in relative terms, the increase in CBWs has recently been much greater than in resident immigrant workers; between 1995 and 2008 (our analysis period), the total number of CBWs increased by about 46.1% (from 146,773 to 214,377); then, between 2008 and 2018, it increased by an additional 46.4% (to a total of 313,926 CBWs).[3] Third and finally, the focus on CBWs gives us the opportunity to implement a simple but powerful instrumental-variable approach based on the country's geography. Obviously, a firm's opportunities to employ CBWs depends on its location relative to the national border. The closer to the border a firm is located, ceteris paribus, the larger the pool of CBWs potentially available for this firm. Thus, we consider the shortest travel time from a firm's location (i.e. the municipality it is located in) to the country border as a valid instrument for the share of CBWs a firm employs, as we will argue in much greater detail below.

Our empirical approach is very close to that used by Dustmann *et al.* (2017), who exploit the introduction of a policy in 1991 that led to a sudden and large influx of Czech CBWs into the German labor market. Specifically, the policy allowed Czech workers to seek employment, but not residence, in border municipalities (see also Moritz, 2011). They find a moderate negative effect on native wages and a rather large negative effect on native employment: a one percentage point increase in the number of Czech CBWs decreased native employment by about 0.93% three years after the policy was implemented. They also find that the employment effects are more pronounced for unskilled workers, who represent a closer substitute for

---

Switzerland, Gerfin and Kaiser (2010) argue that natives and resident immigrant workers are imperfect substitutes, for example.

[3]These figures are taken from the employment statistics of the Swiss Federal Statistical Office, available at https://www.bfs.admin.ch/bfs/en/home/statistics/work-income/surveys/es.html. Note that there are slight discrepancies between these figures and those derived from the data used in the analysis below, due, for example, to different sample definitions.

inflowing Czech workers.

Our empirical strategy is also close to that of Beerli *et al.* (2018), who use a difference-in-differences approach to study how the removal of all remaining immigration restrictions for workers from the European Union increased the number of CBWs to Switzerland and how this in turn affected employment and wages in the regions close to the border. Overall, they find that the removal of the remaining restrictions led to an increase in overall native employment, arguing that the policy increased firms' overall labor demand.

Broadly consistent with this evidence, we find that firms which have easier access to skilled workers partially substitute the training of resident apprentices by hiring additional CBW. Our ordinary least square (OLS henceforth) and two stage least square (2SLS henceforth) estimates are consistent with each other and are robust across a wide range of robustness checks. Moreover, as expected, 2SLS estimates are larger in absolute value than the corresponding OLS estimates, and they are estimated precisely enough to rule out statistical equivalence between OLS and 2SLS estimates. Our preferred 2SLS estimate implies that an increase in the average share of CBWs across firms of 1.14 percentage points, corresponding to the increase observed in the share of CBWs in our analysis period, between 1995 and 2008, leads to a corresponding decrease in the number of apprenticeship positions of about 2%, equal to about 3,500 apprenticeship positions in absolute terms.

In contrast to our study, previous studies on the effects of immigrants on the Swiss labor market have mainly focused on the overall effect of immigration on either wages and/or employment among residents (i.e. natives and previous immigrants).[4] For example, Gerfin and Kaiser (2010), using the structural skill-cell approach, estimate the wage effects from immigration for the period 2002 to 2008. Their findings are mixed: Whereas natives face no wage pressure on average, previous immigrants' wages are negatively affected. Across educational subgroups, highly educated workers incur the largest wage losses. Favre *et al.* (2013) reach similar conclusions,

[4] To save space, we refrain from discussing the voluminous international empirical literature on the effects of immigration. Among many others, both Borjas (2014) and Bansak *et al.* (2015) contain such an overview.

applying a shift-share instrumental-variable approach and focusing on the period between 2002 and 2010. Moreover, their analysis also covers CBWs. They find no overall negative effects of immigrants or CBWs on employment, neither for Swiss-born nor for previously immigrated workers. However, they find statistically significant effects for certain subgroups. A one percent increase in the workforce due to immigration leads to a drop in the employment rate of high skilled Swiss-born workers of 0.3 percentage points. Moreover, whereas immigration does not affect the employment rate of previously immigrated workers, CBW inflows do: a one percent increase in the workforce due to CBWs decreases the employment rate of previously immigrated workers by 0.2 percentage points. Similarly, Basten and Siegenthaler (2019) find that immigration had limited negative effects on natives' employment as well as on their wages. Ruffner and Siegenthaler (2016) explicitly focus on the increase in the number of CBWs in a difference-in-difference approach and conclude that CBWs helped firms overcome their skill shortages and increased firms' productivity without crowding out native employment. The study by Beerli *et al.* (2018) mentioned above also finds that the overall effects of immigration were positive. Our study complements this literature by providing estimates on the substitution of apprentices with CBWs, a question which has not yet been addressed in the empirical literature so far.

The remainder of this chapter is organized as follows. Section 1.2 introduces Switzerland's immigration policy during the analysis period, describes the institutional background of the Swiss apprenticeship system, and formulates our main hypothesis. Section 1.3 presents the data sources and the construction of the key variables, along with descriptive statistics of the key variables. Section 1.4 discusses our empirical strategy. Section 1.5 presents the main results and several robustness checks. Section 1.6 concludes.

1.2 Background

1.2.1 Immigration to Switzerland

Ever since the end of World War II, the proportion of immigrants in the Swiss labor market has been increasing more or less steadily. Starting in the boom of the 1950s, Switzerland became the host country for a large number of immigrants.[5] Consequently, in 2016, the share of non-native residents in Switzerland reached nearly 25%. Initially, and until about the 1980s, immigrants tended to be unskilled laborers (e.g. construction workers), and a majority of them came from geographically close countries (most importantly, from Italy and Spain). Afterwards, the composition of the immigrant inflow changed in two notable ways. First, from about the 1990s onwards, immigration became more heterogeneous; for example, a significant influx of immigrants from former Yugoslavia occurred during and following the Balkan wars. The other notable shift was towards more highly skilled workers, starting in the late 1990s and early 2000s.

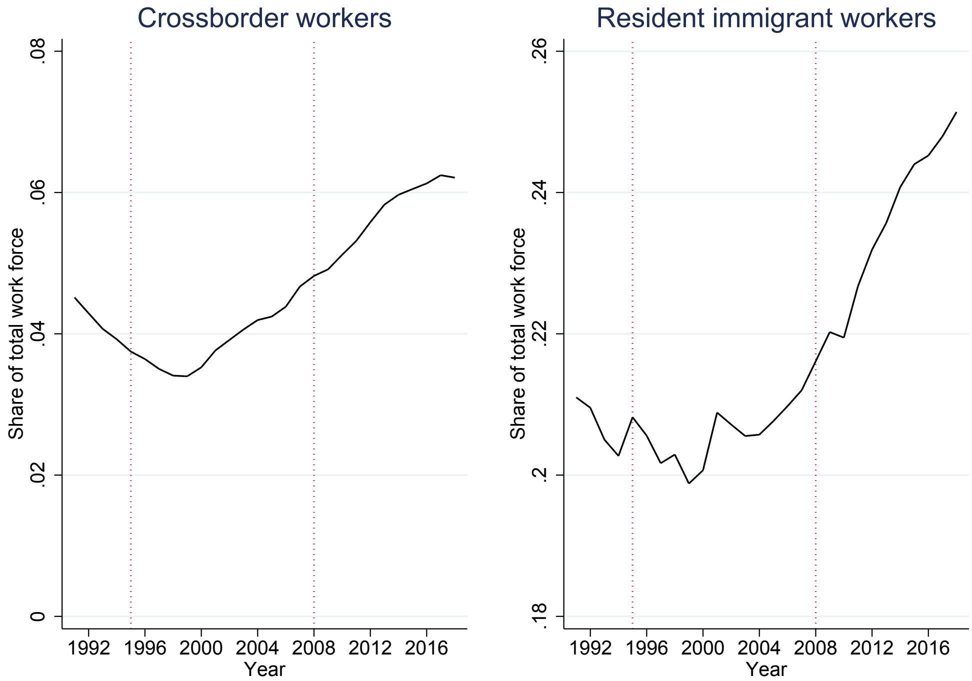
Then, in 1999, Switzerland signed the bilateral Agreement on the Free Movement of Persons (AFMP) with the European Union, which Swiss voters approved in the national plebiscite in 2000. It took effect in 2002 and secured the same legal labor status for both immigrants and CBWs[6] as for native employees. This new system displaced annual quotas, further eased the opportunity for foreign individuals to work within Switzerland, and boosted the numbers both of resident immigrant workers and of CBWs (see figure 1.1).

In 2008, at the end of our analysis period, most of the total of 217,011 CBWs lived in France (119,685), followed by Italy (47,904), Germany

[5]For a long-term overview of Swiss immigration legislation, see Ruedin *et al.* (2015), for example.

[6]In 2002, restrictions on CBWs were first lifted in so-called border regions, which contained municipalities close to the border and were defined in a political process at the state and cantonal levels. The AFMP removed all restrictions on CBWs in border regions in 2004 and in nonborder regions in 2007. However, Ruffner and Siegenthaler (2016) show that the share of CBWs between border and nonborder regions differs only slightly once they control for a firm's distance to the border.

Figure 1.1: Share of CBWs and resident immigrant workers



Notes: The two dotted vertical lines delineate our analysis period (1995 to 2008).

Source: Federal Statistical Office 2017; own calculations.

(42,753), and Austria (6,670). No other country is mentioned in the official statistics on CBWs for 2008. Within Switzerland, 67% percent of all CBWs worked in only three cantons in 2008. The largest group, 54,808 CBWs (25%), worked in the Canton of Geneva, which is located in the far west of Switzerland and almost entirely surrounded by France. In the north of Switzerland, bordering both Germany and France, the two half-cantons of Basel hosted another 22% of the total of CBWs. Italian is the official language in the Canton of Ticino in the south of Switzerland, which due to its proximity to Italy attracted 44,941 CBWs in 2008 (20.7%).[7]

[7]All numbers in this paragraph are taken from the official statistics on cross-border commuters by the Federal Statistical Office; available at https://www.bfs.admin.ch/bfs/en/home/statistics/work-income/surveys/ccs.html. Note that there are slight discrepancies between these figures and those derived from the data

Beside the removal of the remaining legal hurdles in the course of the implementation of the AFMP, there are two main reasons for this large influx of CBWs into the Swiss labor market. First, differentials in real wages are substantial (especially when also considering costs of living, which are considerably lower outside Switzerland) and employment opportunities.[8] Second, the country is landlocked and, due to its small size, the majority of its settlement regions can be reached by car within one hour from a neighboring country (although this ignores the time from a CBW's home to the border). Finally, Switzerland is not a linguistically homogeneous country, nor are its national languages indigenous to the country. In fact, the official language of each of the four countries that border Switzerland is also an official language of Switzerland.[9] Thus, language as a potential remaining barrier for labor market integration after removing legal regulations is basically irrelevant at the Swiss border. Thus, we argue that CBWs' skill mix is relatively similar to that of the resident workforce (cf. Appendix Table B.1).[10]

1.2.2 The Swiss apprenticeship system

In Switzerland, about two thirds of a cohort enter some kind of apprenticeship training after compulsory schooling, which makes apprenticeship training the most important educational track at the upper-secondary level (SERI, 2014). There are presently about 230 occupations to choose from, and regular apprenticeship durations vary between two and four years. In most cases, apprentices sign an apprenticeship contract with a training firm that also pays them a wage, which is however considerably lower than that of a trained

used in the analysis below, due, for example, to different sample definitions).

[8]In fact, there is suggestive evidence that CBWs are paid substantially less than similar native workers (SECO, 2019).

[9]Lichtenstein, a very small country located between Switzerland and Austria, is as good as fully integrated in the Swiss labor and other markets. For this reason, it is considered part of Switzerland in many statistics, including the official statistics on CBWs.

[10]Consequently, mean wages of CBWs (gross monthly median wage of 5,591 CHF) are also more similar to natives' wages (median wage of 5,872 CHF) than those of resident non-native workers (median wage of 4,961 CHF); all wages are computed from the Swiss Earnings Structure Survey of 2004.

worker in the same occupation. The firm is responsible for the practical part of the apprenticeship training. Most apprentices attend vocational school for one or two days a week and spend the rest of the week in their training firms. The alternative, a school-based apprenticeship program, is much less common.[11]

Employers' motives to provide apprenticeship positions: The costs and benefits of apprenticeship training in Switzerland

The Swiss apprenticeship system relies heavily on the participation of mostly private firms, which voluntarily decide whether they want to train apprentices or not. From the employer's point of view, there is a key trade-off between the costs and the benefits of training apprentices (e.g. Mühlemann and Wolter, 2014; Wolter *et al.*, 2006). Training firms bear quite substantial costs, including wages paid to the apprentices and wages for instructors providing the institutionally required on-the-job training for the apprentices. In 2016, these costs averaged roughly 100,000 Swiss francs per apprentice over the full training period, equivalent to about 1.28 annual median wages Gehret *et al.* (2019). However, training firms may profit from the apprentices' productive output during their time spent within the firm; by the end of the training, this output is often not substantially different from that of other employees.

Consistent with this, empirical studies on the costs and benefits of apprenticeship training in Switzerland (e.g. Gehret *et al.*, 2019; Wolter and Strupler, 2012) show that the benefits exceed the costs during the apprenticeship period for about two thirds of all Swiss training firms. Thus, the cumulative benefits from apprentices' productive work by the end of the training period equals or surpasses the cumulative costs of training for these firms. The main reason for this is that apprentices in many occupations become relatively productive early in their apprenticeships, while their wages remain relatively low compared to those of fully trained workers.

[11] See Wettstein *et al.* (2014) for additional background information on the Swiss VET system and how it fits into the country's overall educational system. See also Wolter and Ryan (2011) for a more general discussion of apprenticeship systems.

The costs of hiring skilled workers

At the same time, empirical studies show substantial costs associated with the hiring of skilled workers. In their study focusing specifically on Swiss employers, for example, Blatter *et al.* (2012) document that hiring costs for skilled workers, depending on firm size, equal about 10 to 17 weeks of wage payments on average. Thus an alternative and possibly complementary motivation for training apprentices, especially among those firms incurring net training costs, is the retention of fully trained apprentices by employers to satisfy their demand for skilled labor (e.g. Mohrenweiser and Backes-Gellner, 2010). This motive for training apprentices appears to be especially relevant for firms operating in tight labor markets (Mohrenweiser and Zwick, 2009; Mühlemann and Leiser, 2018).[12]

In line with this argument, Blatter *et al.* (2016) show that Swiss firms that face higher costs for hiring skilled workers from the external labor market tend to train more apprentices and vice versa. They document relatively large effects of hiring costs on the number of training positions, finding that a one-standard-deviation increase in average hiring costs is associated with an increase in the number of apprentices by about half a standard deviation.

1.2.3 Substitution between training apprentices and hiring workers?

Taken together, there are two main motives for training apprentices that are potentially affected by the ease of access to CBWs. First, during their training, apprentices perform productive tasks that otherwise either unskilled or skilled workers perform. Because apprentices earn a much lower wage than either skilled or unskilled workers do, firms may exert a considerable cost advantage when using apprentices for productive work (e.g. Mühlemann and

[12]More recent studies have argued that other motives are relevant as well. For example, firms may use apprenticeship training to screen workers (Mohrenweiser *et al.*, 2017) or as a signaling device (Backes-Gellner and Tuor, 2010). More recently, Kuhn *et al.* (2019) have argued that local norms describing the role of the state also have an influence on firms' decisions to provide apprenticeship positions (cf. chapter 2.

Wolter, 2014). However, this cost advantage of hiring apprentices shrinks if labor becomes more widely available, so wages of skilled or unskilled workers decrease relative to apprentices' wages.

Second, because both the training of apprentices and the hiring of skilled workers from the external labor market are very costly for employers, we expect that easier access to CBWs reduces the hiring costs for employers and may thus tip the balance in favor of hiring workers externally instead of training apprentices.

Overall, both arguments predict partial substitution between training and hiring, and we therefore expect that easier access to CBWs will lead some employers to hire additional CBWs instead of training their own apprentices.

1.3 Data

1.3.1 Firm-level data from the Swiss Business Census

The empirical analysis in this chapter relies mainly on information drawn from the Swiss Business Census (*"Betriebszählung"*). The Business Census covers the population of all firms active in either the second or the third sector in Switzerland, and it includes information on a number of firm-level characteristics, such as the number of employees, industrial affiliation, legal status, geographic location, the number of apprentices, the number of CBWs, and the number of resident immigrant workers (i.e. workers without Swiss citizenship). We have to limit our analysis to the waves of 1995, 2005, and 2008 because only these three include information on the number of CBWs.[13] After 2008, the Business Census was replaced by the firm register dataset STATENT, which no longer includes information on apprentices and is thus unsuitable for our purpose. The availability of data thus restricts the analysis to the period between 1995 and 2008.

The only sample restriction that we additionally impose is that we exclude all firms with fewer than three employees (thus excluding a total of 492,278

[13]However, we do use additional waves of the Business Census to approximate firms' age, as described in more detail below.

observations across the three waves). These firms constitute a significant fraction of all firms (about 43% of all firms in the three waves covered by our analysis), but they cover only a small fraction of all apprenticeship contracts (only about 1.5% of all apprenticeship contracts). Thus, these observations are not really useful when studying firm-provided apprenticeship training, because most of them do not train any apprentices, and we therefore follow the practice of other studies on the subject and exclude these firms from our empirical analysis.[14]

As shown in Table 1.1, our analysis covers a total of 645,137 observations at the firm×year-level, representing a total of 342,323 unique firm observations and 10,595,201 employee observations (see also Appendix Table B.2). In the empirical analysis below, we use the absolute number of apprentices employed by a given firm as our main dependent variable and the share of CBWs as our main regressor (but we will show that these choices are innocuous; see Section 1.5.2 below). Panel (a) of Table 1.1 shows that there are, on average, 0.8 apprentices per firm (however, also note that less than one-third of the firms train any apprentices). Apprentices constitute about 4.8% of the total workforce in our sample in 2008. CBWs represent only about 5.3% of the workforce but about 20% of all non-native workers. Overall, the total number of apprentices in Switzerland grew by 37.3% from 1995 to 2008, while the number of CBWs increased by 50.5%. The total workforce rose by 11.5% in the same period (cf. Table B.2).

Furthermore, we use the Business Census to construct several firm-level controls (descriptives for most of these are shown in panel (b) of Table 1.1). The controls include firms' size by total number of employees and firm size squared, 18 industry dummies, a dummy taking the value 1 if firm i is a private firm and 0 if it is a public firm, and a dummy indicating foreign ownership. Moreover, we add firms' approximate age[15] and firms' share of

[14] For example, in their study on the costs and benefits of apprenticeship training in Switzerland, Wolter and Strupler (2012) also focus on firms with at least three employees. Of course, we will show that our main results are robust with regard to this decision (see Table 1.9).

[15] Specifically, we include a factor variable S_i for the number of sample appearances of

Table 1.1: Descriptive statistics

	Sample Mean	Standard deviation	Minimum	Maximum	Level of aggregation
(a) Key variables					
Number of apprentices	0.775	3.692	0	623	firm
Number of CBWs	0.795	10.592	0	2,503	firm
Share of apprentices	0.059	0.118	0	1	firm
Share of CBWs	0.034	0.116	0	1	firm
Any apprentices (yes = 1)	0.286	0.452	0	1	firm
Any CBWs (yes = 1)	0.131	0.337	0	1	firm
(b) Firm-level controls					
Total number of employees	16.42	66.76	3	10,789	firm
3 - 10 employees	0.710	0.454	0	1	firm
11 - 50 employees	0.239	0.426	0	1	firm
51 - 100 employees	0.030	0.170	0	1	firm
> 100 employees	0.022	0.148	0	1	firm
Share resident immigrant workers	0.170	0.246	0	1	firm
Public firm (yes = 1)	0.116	0.320	0	1	firm
Foreign owned (yes = 1)	0.004	0.065	0	1	firm
Sector III (yes = 1)	0.782	0.413	0	1	firm
(c) Location-specific controls					
Log inhabitants	9.35	1.68	3.09	12.64	municipality
Population density	1,906	2,548	0.740	12,214	municipality
Log firms	8.58	0.902	5.68	10.17	LM-region
Log firms same industry	5.14	1.42	0	8.20	LM-region
% of population working	62.39	2.10	55.61	67.48	LM-region
% Employed in sector III	71.00	10.74	27.08	91.45	LM-region
Median income (CHF)	7,139	691.0	5,500	9,000	LM-region
% Romance	0.290	0.454	0	1	municipality
Social norm	0.196	0.075	0.010	0.956	municipality
(d) Distance to schools in km					
High school	6.60	9.24	0	134.15	municipality
VET school	4.45	6.50	0	70.35	municipality
Full-time VET school	8.78	10.71	0	136.83	municipality
(e) Distance to border					
Travel time by car in minutes	35.38	18.90	3.52	106.26	municipality
Log travel time by car in minutes	3.394	0.634	1.258	4.666	municipality
Observations (year-firm)	645,137	-	-	-	-
Observations (firms)	342,323	-	-	-	-
Observations (municipalities)	2,331	-	-	-	-
Observations (LM-region)	106	-	-	-	-

Sources: Business Census 1995, 2005, 2008; SLFS 2014; Census 2000; Federal Office of Topography 2014; own calculations.

resident immigrant workers to our firm specific controls. The intuition for including these two last control variables is set out in the Appendix A dealing with the validity of the instrument.

1.3.2 Additional regional-level controls

Panels (c) and (d) of Table 1.1 show descriptives for our additional location-specific controls. We also derive the log number of firms within the same labor market region, the log number of firms in the same industry within the same labor market, and the share of employees in the third sector within the respective labor market region from the Business Census. We employ some additional variables from the Federal Statistical Office to control for regional variation in the composition of the residents, such as the logarithm of municipalities' inhabitants and their population density. Furthermore, we calculate the median income for every labor market region from the Swiss Labor Force Survey 2014 (*"Schweizerische Arbeitskräfteerhebung"*) to control for regional income differentials. Finally, we include 26 cantonal and three language dummies. Note that the regional variables are measured at either the municipal level or at the level of 106 labor-market regions (cf. last column of Table 1.1).

Moreover, following the analysis by Kuhn *et al.* (2019), we also control for a local norm favoring the private over state provision of public goods and consequently influencing firms' investment in apprenticeship training. Controlling for this norm could be important in our context because there is a pronounced regional pattern in the strength of the norm. We follow Kuhn *et al.* (2019) and use the municipality results from two national-level votes on the role of the state within VET policy to measure this norm.[16]

firm i, e.g. for firm i we do not observe in the waves 1991 and 1995 but in the waves 1998 to 2008, $S_i = 3$ in 2005 and $S_i = 4$ in 2008. Note that waves 1998 and 2001 are not part of our sample because they lack information on CBWs; however, we observe whether a firm existed in these waves.

[16] We use the mean share of supporting votes from two popular initiatives that demanded a stronger role of the state in the VET system (i.e. smaller values on this variable denote a stronger norm towards the private provision of training). See Kuhn *et al.* (2019) for

Although the principal aim of this chapter is to analyze the effect of CBWs on firms' supply of apprenticeships, note that we observe concluded apprenticeship contracts, the number of which is likely affected by pupils' demand for apprenticeships as well.[17] For this reason, we also include a firm's distance from the nearest high school, VET school, and full-time VET school as exogenous demand shifters in most of our estimations.

1.3.3 Distance to the national border

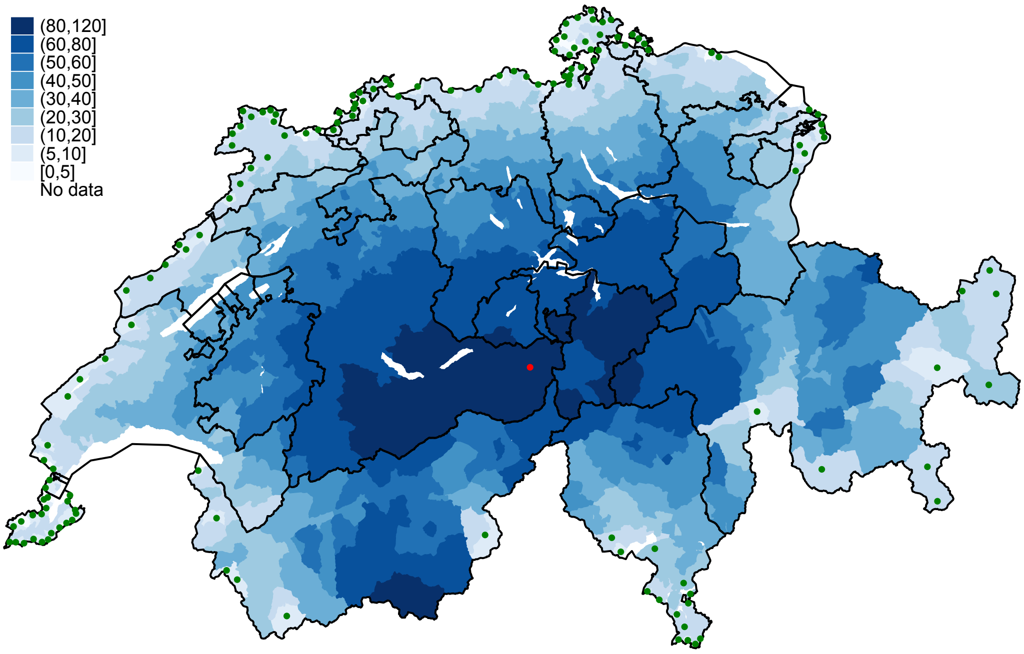
A final variable of interest is a firm's distance from the national border, which we will use as instrument for the share of CBWs within a firm (this idea is discussed in Section 1.4 below in detail). To construct this variable, we use the matrix of travelling times put together by the Federal Office of Spatial Development, which contains the average travel time by car in minutes between any two Swiss municipalities (it also contains average travelling time when moving from one place to another within the same municipality). We set the minimum distance from the national border for every firm equal to the minimum travelling time by car from the municipality within which a given firm is located to one of the 138 different municipalities with a navigable border crossing. See Figure 1.2 for a graphical representation of this variable.

On average, firms included in our analysis lie 36.7 minutes by car away from the border. There is large variation in the travelling time to the border: The minimum travel time is only 3.5 minutes (for the municipality of *Le Grand-Saconnex* in the Canton of Geneva), whereas the maximum is 106.3 minutes (for the municipality of *Innertkirchen* in the Canton of Bern; cf. Figure 1.2). In the analysis below, we use the natural logarithm of this variable, which ranges from 1.26 to 4.67 (see panel (e) of Table 1.1 for the corresponding descriptives).

additional details.

[17]Note in this context that we deviate from the standard notation in labor economics as we distinguish between firms' supply of and pupils demand for apprenticeships, whereas in labor economics firms usually demand labor and workers supply it. This also applies in Chapter 2.

Figure 1.2: Travelling distance from the national border



Notes: The figure shows the minimum travelling distance from each municipality to the national border. The green dots indicate the center of a bordering municipality; the red dot indicates the center of *Innertkirchen*, the municipality most far away from the national border.

1.4 Empirical strategy

1.4.1 OLS estimates

To estimate the effect of CBWs on firms' training incidence, we start with the following basic regression equation:

$$A_i = \alpha + \beta CBW_i + \gamma x_i + \delta z_{j[i]} + \psi_{t[i]} + \epsilon_i \tag{1.1}$$

The dependent variable A_i corresponds to the absolute number of apprentices employed by firm i, and the regressor of main interest is the share of CBWs working in firm i, denoted by CBW_i in Equation (1.1).[18] Most specifications

[18]More specifically, the share of CBWs is the ratio of the number of CBWs to the total number of employees including apprentices. Because apprentices are different from fully-

will control for several firm-level characteristics described in Section 1.3.1, denoted by x_i. The characteristics of region j (see Section 1.3.2), where firm i is located in, enter as $z_{j[i]}$ on the right-hand side of Equation (1.1). Finally, we also include a full set of census-year fixed effects, denoted by $\psi_{t[i]}$.

Whatever the exact set of controls, β is the parameter of main interest, because it quantifies the partial association between a firm's share of CBWs and its number of apprentices. Throughout the analysis, we present robust standard errors that are clustered at the municipality level to account for potential correlation in the error terms of firms located within the same municipality (e.g. Cameron and Miller, 2015).

1.4.2 Instrumental-variable estimates

However, even after controlling for a large set of firm- and regional-level controls, OLS estimates of the parameter β from Equation (1.1) may suffer from endogeneity bias. Assume, for example, that firms hire additional CBWs to satisfy an increasing demand for their products during an economic upswing. For the same reason, firms concurrently hire more apprentices performing productive tasks. In this case, OLS estimates would reveal a positive correlation between firms' share of CBWs and the number of apprentices they employ, though no causal interpretation applies.

To account for potential endogeneity issues of this kind, we propose an IV approach exploiting the fact that commuting distances impose considerable costs on CBWs, and distance to the border therefore constrains firms' hiring of CBWs.[19] Consequently, we instrument the share of CBWs employed by firm i with the minimum distance to the national border of municipality j firm i is located, measured by the log travel time by car, $\ln(D_{j[i]})$.[20] Thus,

trained workers, and because they implicitly appear on both sides of Equation (1.1), we also checked that our results are robust to an alternative construction of this variable not counting apprentices as part of a firm's workforce (see Section 1.5.2 below).

[19]More specifically, commuting costs consist of both direct transportation costs and the opportunity costs arising during the commuting time.

[20]We chose this parameterization of the instrument because it yields the highest fit in the first-stage regression. We also checked that other parameterizations yield similar

we argue that variation in municipalities' minimum car-driving time to the national border induces quasi-experimental variation in firms' possibility to employ CBWs. The following two equations capture the essence of this approach:

$$A_i = \alpha + \beta \widehat{CBW}_i + \gamma x_i + \delta z_{j[i]} + \psi_{t[i]} + \epsilon_i, \tag{1.2}$$

where A_i, x_i, $z_{j[i]}$, and $\psi_{t[i]}$ are the same as in Equation (1.1) discussed above, but where $\widehat{CBW}_i$ comes from the following first-stage regression:

$$CBW_i = \pi_0 + \pi_1 \ln(D_{j[i]}) + \pi_2 x_i + \pi_3 z_{j[i]} + \psi_{t[i]} + \upsilon_i. \tag{1.3}$$

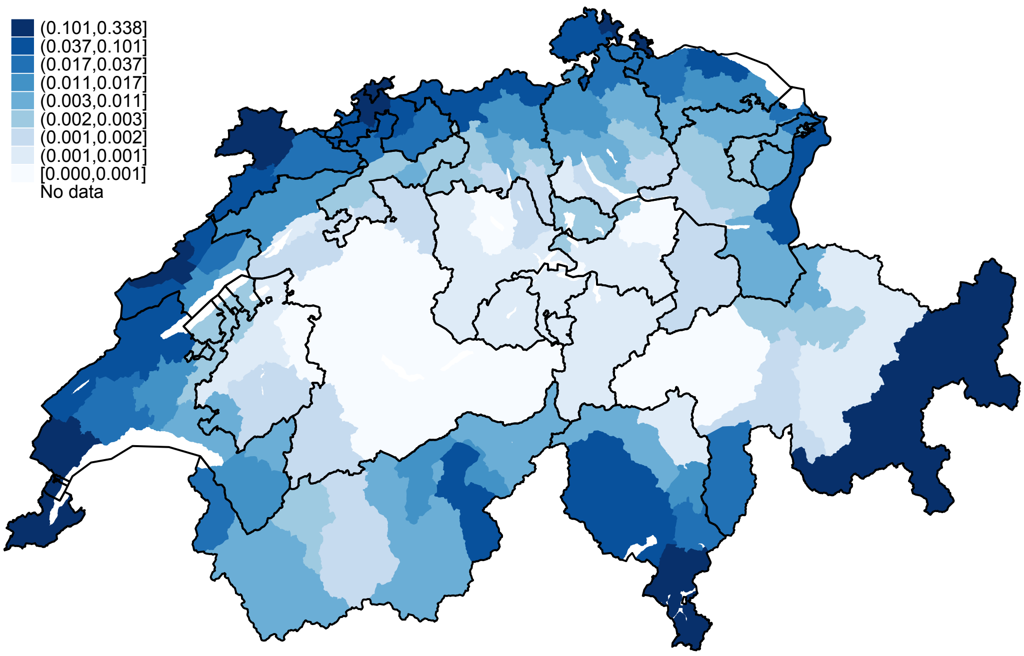
As usual in such a setup, we use Equation (1.3) to empirically test the strength of the instrument (first-stage estimates, and associated test statistics, are discussed in Section 1.5.1). The instrument should also meet the additional requirements of being as good as randomly assigned (possibly conditional on some set of controls) and not having a direct effect on the outcome (e.g. Angrist and Pischke, 2008; Wooldridge, 2010). We discuss these two additional requirements, which cannot directly be tested empirically, in Appendix A, and we present some robustness checks related to the concern of endogenous locational decisions of employers in Section 1.5.2.

Figure 1.3 shows the spatial distribution of the share of CBWs at the district level. Supporting our argument above, it is immediately evident that the density of CBWs decreases strongly with a region's distance from the national border. As a consequence, variation among the 148 districts is large and ranges from the district of *Gersau* in in the Canton of Lucerne in central Switzerland with no CBWs at all to *Mendrisio* in the Canton of Ticino with a share of CBWs of 34 percent.

---

results (see Table 1.5.2 below).

Figure 1.3: Share of CBWs at districct level



Notes: The figure shows the spatial distribution of the share of CBWs (i.e. the ratio of the number of CBWs to the total work force) at the district level (there are 148 distinct districts).

1.5 Results

1.5.1 Main results

OLS estimates

Table 1.2 shows OLS estimates of a series of regressions that relate a firm's number of apprentices to its share of CBWs, according to the setup of equation (1.1).

First, the raw association between the number of apprentices and the share of CBWs, shown in column (1) of Table 1.2, is negative but not statistically significant ($\widehat{\beta} = -0.220$). The next column adds census-year dummies, leaving the relevant estimate virtually unchanged ($\widehat{\beta} = -0.250$). Column (3) adds firm-level controls, which yield a negative and statistically significant estimate ($\widehat{\beta} = -1.039$). Thus, compared to the previous two

Table 1.2: Main results (OLS estimates)

	Number of apprentices				
	(1)	(2)	(3)	(4)	(5)
Share of CBWs	-0.220	-0.250	-1.039***	-0.710***	-0.709***
	(0.171)	(0.173)	(0.114)	(0.133)	(0.133)
	[-0.010]	[-0.011]	[-0.045]	[-0.031]	[-0.031]
Census year dummies	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Firm controls	*No*	*No*	*Yes*	*Yes*	*Yes*
Location controls	*No*	*No*	*No*	*Yes*	*Yes*
Distance to schools	*No*	*No*	*No*	*No*	*Yes*
R-squared	0.000	0.001	0.298	0.300	0.300
Observations	645,137	645,137	645,137	645,137	645,137

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

columns, this specification points to a possible substitution by of training apprentices with hiring CBWs. The change in the estimated coefficient also suggests that those firms that are, ceteris paribus, more likely to hire apprentices are also more likely to hire CBWs. Further adding regional controls, including the full set of cantonal dummies, yields a comparable, yet somewhat smaller estimate of $\widehat{\beta} = -0.710$, suggesting that some of the regional-level characteristics are correlated with both the incidence of training and the likelihood of hiring CBWs. More precisely, the change in $\widehat{\beta}$ suggests that regional features that make firms in these locations more likely to train apprentices are negatively correlated with the share of CBWs. Finally, we add the three distance-to-schools variables in the fifth, which again yields an estimate very close to that in the previous column. The estimate of $\widehat{\beta} = -0.709$ in the fifth column implies that a 10 percentage point increase in employers' share of CBWs is associated with 0.071 fewer apprentices per employer on average (we provide a more detailed discussion of the economic size of the estimated effects further below).

2SLS estimates

We next turn to the instrumental variable approach introduced in Section 1.4.2, using a firm's distance to the national border as an instrument for its share of CBWs.[21]

Table 1.3: Main results (2SLS estimates)

	Number of apprentices				
	(1)	(2)	(3)	(4)	(5)
Share of CBWs	-1.778***	-1.790***	-1.958***	-1.319***	-1.445***
	(0.437)	(0.438)	(0.301)	(0.291)	(0.272)
	[-0.077]	[-0.077]	[-0.085]	[-0.057]	[-0.062]
Census year dummies	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Firm controls	*No*	*No*	*Yes*	*Yes*	*Yes*
Location controls	*No*	*No*	*No*	*Yes*	*Yes*
Distance to schools	*No*	*No*	*No*	*No*	*Yes*
R-squared	0.000	0.000	0.298	0.300	0.300
Observations	645,137	645,137	645,137	645,137	645,137
F-statistic (first-stage)	201.6	201.8	174.4	170.7	183.0

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

Table 1.3 presents the resulting 2SLS estimates and also starts with a simple specification that includes no control variables at all. Instrumenting a firm's share of CBWs with its minimum distance to the border yields a negative and statistically significant estimate ($\widehat{\beta} = -1.778$). Moreover,

[21] Appendix Tables B.3 and B.4 show the corresponding first-stage and reduced-form estimates, respectively. Both effects are statistically significant and, more importantly, sizable and rather precisely estimated. Also note that both are relatively stable across the various specifications.

in comparison with the corresponding OLS estimate, the 2SLS estimate is considerably larger, and the increase in the estimate also makes up for the parallel increase in the standard error relative to the OLS estimate.[22]

We include census-year fixed effects in column (2) to account for changes in Switzerland's immigration policy during our sample period. However, this hardly changes the relevant estimate (mirroring the pattern from Table 1.2 above), which equals $\widehat{\beta} = -1.790$ in this specification.[23] In column (3), we add our full set of firm-specific controls. The resulting estimate of $\widehat{\beta} = -1.958$ is somewhat smaller, but still close to that in the previous column, and it remains highly statistically significant. Column (4) of Table 1.3 adds our set of regional controls. This yields a somewhat smaller (i.e. less negative) estimate of $\widehat{\beta} = -1.319$, which, however, remains highly significant.

Finally, the specification shown in column (5) also includes a firms' minimum distance to three schooling facilities (dual VET school, full school VET institutions, and high school). This specification, including the full set of controls, yields an estimate of $\widehat{\beta} = -1.445$, which remains highly statistically significant. This is our preferred estimate of the effect of CBWs on firms' provision of apprenticeship positions. Compared to the corresponding OLS estimate from column (5) of Table 1.2, the 2SLS estimate is about twice as large (i.e. more negative). Moreover, note that both OLS and 2SLS estimates are quite precisely estimated (the OLS and 2SLS estimates have similar t-values).

Quantifying the economic size of the estimated effect

Taken together, the estimates shown in Tables 1.2 and 1.3 support our hypothesis that firms tend to substitute between training resident apprentices and hiring CBWs when access to CBWs is easier. Also consistent with our

[22]Consistent with this estimate, Appendix Figure B.1 shows that firms located in regions closer to the national border tend to train fewer apprentices on average than firms in regions further away from the border.

[23]We also estimated a model that included the interaction terms between the census-year dummies and the share of CBWs. The resulting 2SLS estimates on the interaction terms are positive, but statistically not significantly different from zero.

prior expectations, 2SLS estimates are larger (i.e. more negative) than the corresponding OLS estimates. Before providing additional robustness checks, we discuss the economic size of the estimates.

Some back-of-the-envelope calculations may illustrate the economic size of the estimated effect. Our preferred 2SLS estimate (i.e. that from column (5) of Table 1.3) of $\widehat{\beta} = -1.445$ implies that a ten percentage point increase in the mean share of CBWs decreases the number of apprentices trained by employers by 0.1446 on average. However, most firms presumably employ much smaller shares of CBWs (see Table 1.1 again). Thus, as a more realistic benchmark, we may use the change observed in the share of CBWs between 1995 and 2008, the time period covered by our analysis. Within this time period, the observed share of CBWs for the average firm increased from 2.73% to 3.87%: an increase of 1.14 percentage points (cf. Table B.2).[24]

Everything else constant, our baseline 2SLS estimate thus informs us that this will translate into a decrease of 0.0164 apprentices by each employer. There were 212,081 firms in our sample in 1995, and 219,950 in 2008. Thus, approximately 3,500 apprenticeship positions were substituted by CBWs within this time period (-3,493 to -3,629, depending on whether we use the absolute number of firms in 1995 or in 2008). This corresponds to about 2% of the total of apprenticeship positions in relative terms (-2.498% to -1.891%).

[24] Note that the share of CBWs at the level of the firm differs from the share of CBWs in the workforce (compare the statistics from Tables 1.1 and B.2). Apparently, larger firms tend to hire more CBWs than smaller firms.

1.5.2 Robustness

We next probe the robustness of our estimates.

Alternative specifications

As a first check, Table 1.4 presents a series of alternative specifications of our baseline model from column (5) of Table 1.3 (note that the first column of Table 1.4 simply replicates this specification for the ease of comparison). We estimate specifications that use a slightly different parameterization of the instrument. Instead of using $\ln(D_{j[i]})$ as in the baseline specification, we use either $D_{j[i]}$, a quadratic in $\ln(D_{j[i]})$ or a quadratic in $D_{j[i]}$ (columns (2) to (4) in Table 1.4). Evidently, all these alternative specifications yield estimates of β close to our baseline estimate; in these three specifications, $\widehat{\beta}$ varies between -1.637 and -1.086.

The remaining four columns use different parameterizations of either the endogenous variable or the dependent variable. Specifically, we use the absolute number of CBWs (instead of the share of CBWs) as endogenous variable in column (5). By construction, this yields a somewhat different estimate of $\widehat{\beta} = -0.068$; however, note that the associated elasticity is similar to that in column (1), -0.07 versus -0.062. Another variation, shown in column (6) of Table 1.4, uses a slightly different definition of the share of CBWs. While we do count apprentices as employees in the baseline specification, we construct the share of CBWs in this specification as the ratio of the number of CBWs to the number of employees excluding apprentices. Again, this yields an almost identical estimate ($\widehat{\beta} = -1.392$) as our baseline specification. In column (7), we regress the share of apprentices on the share of CBWs. The resulting estimate of $\widehat{\beta} = -0.110$ is substantially smaller than that from the baseline specification. Note, however, that this is primarily due to the difference in the scaling of the dependent variable (also note that, again, the implied elasticity is virtually identical to that from the baseline specification). Finally, we construct a dummy being 1 for training firms and 0 otherwise in column (8), which we then regress on the share of CBWs. The estimate ($\widehat{\beta} = -0.336$) yields the substitution effect between CBWs and

Table 1.4: Alternative specifications (2SLS estimates)

	Number of apprentices						Share of apprentices	Training dummy
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share of CBWs	-1.445***	-1.637***	-1.086***	-1.456***			-0.110***	-0.336***
	(0.272)	(0.450)	(0.288)	(0.247)			(0.014)	(0.051)
	[-0.062]	[-0.071]	[-0.047]	[-0.063]			[-0.063]	[-0.039]
Number of CBWs					-0.068***			
					(0.013)			
					[-0.070]			
Share of CBWs^a						-1.392***		
						(0.263)		
						[-0.063]		
Census-year dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Location controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance to schools	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
R-squared	0.300	0.300	0.300	0.300	0.277	0.299	0.088	0.117
Observations	645,137	64,5137	645,137	645,137	645,137	645,137	645,137	645,137
F-statistic (first-stage)	183.0	104.2	89.1	92.6	81.9	188.9	183.0	183.0

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities. Instead of (only) $\ln(D_{j[i]})$, we include $D_{j[i]}$ in column (2), $\ln(D_{j[i]})$ and a quadratic in $\ln(D_{j[i]})$ in column (3), and $D_{j[i]}$ and a quadratic in $D_{j[i]}$ in column (4). ^{a}Refers to a firm's ratio of CBWs to total employees excluding apprentices.

apprentices on the extensive margin. Thus, easier access to CBWs leads some firms to stop training apprentices altogether.

Overall, our estimate of the substitution between CBWs and apprentices appears robust to some obvious alternative parameterizations of the instrument and to the endogenous and the dependent variables. We next turn to some more subtle issues.

Unobservable firm characteristics

In our baseline estimates (Table 1.3) we include various firm-specific controls in an effort to account for the fact that the potential to hire apprentices varies substantially among firms, for example of different size or across different industries. However, other firm characteristics interacting with firms' provision of apprenticeship positions might be unobservable.[25] To dispel concerns to that effect, Table 1.5 applies our estimation approach to differenced data in a subsample of 212,730 firms surveyed at least twice within our sample.[26]

The regressions underlying the estimates shown in Table 1.5 deviate slightly from the conventional instrumental variable setting using difference data because our instrument does not vary at all over time. We therefore instrument the change in the share of CBWs with a firm's distance to the border as well. This implies that we assume that the distance to the border not only predicts the share of CBWs in a given year but also the change in the share of CBWs between two consecutive waves of the business census.[27] The

[25] In this context, Beerli *et al.* (2018) present evidence that the inflow of CBWs affected firms in many dimensions, e.g. increasing their size and productivity. Exploiting within-firm variation in the share of CBWs in this section also addresses concerns that such changes in firm characteristics threat our identification strategy.

[26] The total of 332,456 observations in Table 1.5 consist of 20,246 observations from firms observed in 1995 and 2005, 72,758 observations from firms observed in 2005 and 2008, and 239,452 observations from 119,726 firms observed in all three waves.

[27] That is, we estimate regressions of the following form:

$$\Delta A_i = \beta \Delta CBW_i + \gamma \Delta x_i + \delta \Delta z_{j[i]} + \Delta \epsilon_i,$$

corresponding first-stage F-statistics in Table 1.5 displaying values between 45.28 and 52.60 confirm this.

Table 1.5: First-differenced data (2SLS estimates)

	Δ Number of apprentices		
	(1)	(2)	(3)
Δ Share of CBWs	-1.572	-3.000**	-3.006**
	(1.215)	(1.189)	(1.195)
	[-0.045]	[-0.086]	[-0.088]
Δ Firm controls	*No*	*Yes*	*Yes*
Δ Location controls	*No*	*No*	*Yes*
R-squared	0.000	0.066	0.066
Observations	332,456	332,456	332,456
F-statistic (first-stage)	45.28	51.18	52.56

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

The estimate in column (1) of Table 1.5 has about the same size as the baseline estimate from Table 1.3 but is statistically insignificant due to its large associated standard error. In contrast, the estimates in columns (2) and (3) of Table 1.5 suggest that an increase in the share of CBWs leads to substitution away from apprentices. Taking into account that these estimates are relatively imprecisely estimated, they appear in line with the baseline estimate from Section 1.5.1. This in turn suggests that our main finding remains robust when accounting for unobservable firm characteristics.

Firms' age and endogenous locational choices by employers

Endogenous location decisions by firms pose a potential threat to our identification strategy. We discuss this issue in some detail in Appendix A

where Δ denotes that a variable has been differenced. As in the main analysis, we instrument ΔCBW_i with $\ln(D_{j[i]})$.

and argue that the evidence points towards relatively large differences in the provision of apprenticeship positions between different types of firms. For example, newly established firms, as well as firms changing their location, tend to train fewer apprentices than existing and non-moving firms.[28] At the same time, these differences in training behavior also appear to correlate with locational choices of employers, even though the differences in distance from the border are relatively small. In combination, these two findings may imply that endogenous location decisions could impact our 2SLS estimates.

Table 1.6 first tackles concerns regarding endogenous location decisions by estimating our baseline specification within various subsamples of firms of similar age; consequently, the time at which and the duration since these firms made their initial location choices are comparable. In order to assign firms to any of the three subsamples of columns (2) to (4) in Table 1.6, we approximate their age by identifying firms throughout our sample period. Concretely, we identify three types of firms. Firms included in all waves enter the subsample of column (2). These firms were already established in 1991 (first wave) and remained within our sample until 2008. The approximation of firms' age within this subsample remains relatively vague, because we can only state that these firms were present in our sample for the same duration, but we are unable to estimate the duration of their existence prior to 1991. The column (3) subsample of newly established firms only considers observations of firms appearing in our sample for the first time; hence, these firms are between zero and three years (the maximum number of years between two waves) old. In column (4), we construct a subsample of middle-aged firms by restricting the sample to observations of the 2005 and 2008 waves, respectively, and only include firms which appear for the first time in our sample either in 1995 or 1998. This limits the subsample in column (4) to firms of age seven to sixteen years.[29] All estimates in Table 1.6

[28]However, the finding that newly established firms train fewer apprentices is not new (see Müller and Schweri, 2012, for example).

[29]Among these firms, the youngest possible firms appeared for the first time in the 1998 wave (founded in 1998) and are observed in 2005. The possibly oldest firms (founded in 1992) appear for the first time in the 1995 wave and are observed in 2008.

Table 1.6: Heterogeneity by firms' age (2SLS estimates)

	Number of apprentices				
	(1) Baseline (all firms)	(2) Included in all waves	(3) Newly established	(4) "Middle aged" firms	(5) Moving firms only
Share of CBWs	-1.445***	-1.921***	-0.611**	-1.312**	-1.204*
	(0.272)	(0.408)	(0.269)	(0.516)	(0.701)
	[-0.062]	[-0.052]	[-0.069]	[-0.071]	[-0.145]
Census year dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Location controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance to schools	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
R-squared	0.300	0.398	0.090	0.156	0.091
Observations	645,137	292,838	90,238	56,194	24,943
F-statistic (first-stage)	183.0	202.4	127.5	123.4	102.8

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

are statistically significant. The size of the estimates in column (2) and (4) are fairly similar to our baseline specification in column (1), whereas the estimate in column (3) is somewhat smaller. Given the lower number of apprentices across newly established firms in the subsample of column (3), this smaller estimate seems plausible and corresponds to a similar effect in relative terms.[30]

Finally, column (5) focuses on the small subsample of firms which move to another municipality between two consecutive waves of the business census (see also Appendix A for additional details). We find $\widehat{\beta}$ to be of similar size to the baseline estimate when using only this subsample of moving firms. Overall, we conclude that the negative effect of CBW inflows on firms' training provision pertains across firms of various ages and is therefore neither biased by the establishment of new firms nor the moving of existing firms.

Additional robustness checks

Table 1.7 presents some additional robustness checks. Again, the first column of Table 1.7 simply replicates the baseline specification column (5) of Table 1.3. In column (2) of Table 1.7, we first enlarge the sample by also including 492,278 observations from firms with 2 or fewer employees (cf. Table B.2). The resulting 2SLS estimate of $\widehat{\beta} = -1.065$ becomes less negative than our baseline specification because firms' average number of apprentices among this subsample is smaller. However, considering the statistical imprecision, the estimate in column (5) is again close to our baseline specification.

Column (3) of Table 1.7 limits our sample to 256,040 firms located within 30 minutes' car driving from the border (i.e. $D_{j[i]} \leq 30$ or $\ln(D_{j[i]}) \leq 3.40$, respectively). We argue that regions within this small bandwidth are comparable, e.g. in geography or economic structure. Reassuringly, the

[30]The average number of apprentices per firm among the subsamples of columns (1) to (4) are 0.78, 1.05, 0.33, and 0.66. Multiplying the estimates with a 10 percentage point increase in the share of CBWs yields a substitution of 18% to 20% of the total number of apprentices per firm across all subsamples of Table 1.6.

Table 1.7: Additional robustness checks (2SLS estimates)

	Number of apprentices							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline	Including micro firms	$D_{j[i]}$ < 30	For-profit firms	Within MS-region	Wave 1995	Wave 2005	Wave 2008
Share of CBWs	-1.445***	-1.065***	-1.828***	-1.363***	-1.688***	-1.373***	-1.496***	-1.501***
	(0.272)	(0.227)	(0.364)	(0.251)	(0.478)	(0.266)	(0.318)	(0.316)
	[-0.062]	[-0.059]	[-0.198]	[-0.065]	[-0.073]	[-0.057]	[-0.064]	[-0.068]
Census year dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Location controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance to schools	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
R-squared	0.300	0.306	0.261	0.292	0.300	0.303	0.283	0.314
Observations	645,137	1,136,188	256,040	570,606	645,137	212,081	213,106	219,950
F-statistic (first-stage)	183.0	158.4	95.5	185.6	99.7	217.7	177.0	155.6

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

estimate of $\widehat{\beta} = -1.828$ applied for firms located close to the border is again of similar magnitude as in our baseline specification.

Next, we limit our sample to private for-profit firms in column (4). The coefficient of $\widehat{\beta} = -1.363$ in column (4) pertaining for for-profit firms is comparable to the baseline estimate in column (1). This assures us that our results are not driven by state authorities, which arguably face different budget restrictions than privately-run firms.

In column (5) we control for a full set of local labor-market fixed effects (using a set of 106 dummies) instead of including cantonal-level dummies. While cantons are Switzerland's most important jurisdictional entities, labor markets are often not bound to their borders and are alternatively defined by the Federal Statistical Office by commuting patterns. The resulting estimate of $\widehat{\beta} = -1.688$ in column (5) thus yields the effect of CBW inflows on the provision of apprenticeship positions by firms that operate in the same market in terms of product demand and labor supply.

In the last three columns of Table 1.7, we check whether the negative effect of the share of CBWs on firms' training provision is persistent throughout all three waves of the Business Census covered by our sample. The three corresponding estimates of columns (6) to (8) vary little (between $\widehat{\beta} = -1.373$ to $\widehat{\beta} = -1.501$), their confidence intervals overlap widely, and they are almost identical to the estimate in the baseline specification of column (1).

1.5.3 Heterogeneity

In the final part of the empirical analysis, we check whether we find heterogeneous effects conforming to prior expectations to further corroborate our main finding.

Training costs

In this subsection, we test whether varying apprenticeship costs among firms translates into heterogeneous substitution effects between CBWs and apprentices. A priori, we expect the substitution effect to increase with firms' apprenticeship costs for two reasons. First, roughly two thirds of all

Swiss training firms profit financially from training apprentices.[31] Arguably, the incentive to substitute apprentices with CBWs is relatively low for firms that profit financially from their apprentices. Second, firms bearing net costs during an apprenticeship often seek to satisfy their future skill demand by retaining fully trained apprentices (Blatter *et al.*, 2016). Inflows of CBWs, which enlarge the supply of skilled workers and increase the possibility to recruit on the external labor market, lower this particular incentive of firms to provide apprenticeship positions.

We test this hypothesis by first estimating β across subsamples defined by the average net cost of an apprenticeship $\overline{c}_{k[i]}$ in industry k firm i operates in (see Table 1.8, columns (1) to (3)).[32] Whereas the first two columns aggregate firms operating in industries that display either strongly negative (column 1) or moderately negative (column 2) net costs (i.e. net profits in these cases) from apprenticeship training, column (3) aggregates industries in which firms on average bear positive net costs from offering apprenticeship positions.

The estimate of $\widehat{\beta} = -0.580$ in column (1) is relatively small and statistically insignificant. This suggests that firms that tend to profit financially from training apprentices are less likely to substitute apprentices with CBWs. In contrast, the estimate from column (2) suggests a negative association between the share of CBWs and the number of apprentices. Apparently, these firms partly substitute apprentices with

[31] This applies especially in occupations where training curricula require few instructions and little time for practicing. In these occupations, the productivity gap between apprentices and regular workers is therefore relatively low. Another third of the Swiss training firms bear sometimes substantial net cost during an apprenticeship; see Gehret *et al.* (2019) and Section 1.2.2 for details.

[32] Ideally, we would like to calculate the apprenticeship net costs at firm level or at least occupational level because training curricula, which partly determine firms' VET costs, are occupation specific. Unfortunately, we do not observe occupations in the Business Census. As a second-best solution, we approximate firms' apprenticeship costs by the average net costs within 82 industries (at the NOGA 2-digit level) derived from the most recent data on the costs and benefits of apprenticeship training available (Gehret *et al.*, 2019).

Table 1.8: Heterogeneity by costs of apprenticeship training (2SLS estimates)

	Number of apprentices			
	(1)	(2)	(3)	(4)
$\overline{c}_{k[i]}$	$< p25$	$\in [p25, p75]$	$> p75$	Full sample
Share of CBWs	-0.580	-1.257***	-3.344***	-1.724***
	(0.393)	(0.303)	(0.726)	(0.285)
	[-0.022]	[-0.057]	[-0.150]	[-0.074]
Share of CBWs$\times\overline{c}_{k[i]}$				-0.042***
				(0.010)
$\overline{c}_{k[i]}$				-0.003***
				(0.001)
Census year dummies	*Yes*	*Yes*	*Yes*	*Yes*
Firm controls	*Yes*	*Yes*	*Yes*	*Yes*
Location controls	*Yes*	*Yes*	*Yes*	*Yes*
Distance to schools	*Yes*	*Yes*	*Yes*	*Yes*
R-squared	0.333	0.243	0.462	0.304
Observations[a]	159,274	373,078	111,988	644,340
F-statistic (first-stage) for:				
Share of CBWs	180.0	144.2	210.5	95.16
Share of CBWs $\times\overline{c}_{k[i]}$				131.15

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities. $\overline{c}_{k[i]}$ denotes the average costs of an apprenticeship in the NOGA-2-digit category k firm i operates in. $p25$ and $p75$ denote the 25th and 75th percentile of the distribution of $\overline{c}_{k[i]}$, respectively. [a]Because we do not observe at least one apprenticeship in every NOGA-2-digit industry in the cost/benefit data, we lose some observations compared to our baseline sample.

CBWs even though they also tend to make a profit from training. One potential explanation for this pattern might be that firms usually recruit teenagers with no previous work experience at all for open training positions. This considerably limits firms' ability to assess apprentices' productivity. Therefore, the employer is often very uncertain about the expected costs and benefits. The fact that about 20% of all apprenticeship contracts end prematurely further adds to this uncertainty. In contrast, CBWs with a formal degree and positive work experience likely represent a more predictable alternative to firms. Column (3) focuses on industries in which firms bear net costs from offering an apprenticeship on average. The estimate of $\widehat{\beta} = -3.344$ reveals that the rate of substitution is much higher in this particular subsample of firms.

Finally, we return to the full sample in column (4) but simply interact the share of CBWs, and consequently also the instrument $\ln(D_{j[i]})$, with the industry average costs of an apprenticeship $\overline{c}_{k[i]}$. Consistent with the estimates from columns (1) to (3), the negative coefficient associated with the interaction term implies that the effect of CBWs on firms' VET provision increases with respect to apprenticeship net costs of industry k in which firm i is operating.

Additional subsample results

As a final robustness check, we provide several additional subsample results in Table 1.9 and compare them to the baseline specification displayed in column (1) for ease of comparison.

First, columns (2) and (3) of Table 1.9 present subsample estimates among small (50 or less employees) and large (more than 50 employees) firms, respectively. Both estimates are statistically significant at the 1% level, but $\widehat{\beta} = -11.523$ in column (3) appears much larger than $\widehat{\beta} = -0.874$ in column (2). These differences arise largely due to different levels of the outcome variable across the two subsamples, however. In the baseline sample of column (1), firms employ an average of 0.78 apprentices, while this number amounts to 5.28 among firms with 50 or more employees considered in column

Table 1.9: Subsample results (2SLS estimates)

	Number of apprentices						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Baseline	Small firms	Large firms	German language	Romance language	BL, BS, GE, TI excluded	BL, BS, GE, TI only
Share of CBWs	-1.445***	-0.874***	-11.523***	-1.351**	-0.883***	-1.412**	-1.342***
	(0.272)	(0.130)	(2.364)	(0.528)	(0.256)	(0.406)	(0.443)
	[-0.062]	[-0.053]	[-0.119]	[-0.026]	[-0.122]	[-0.023]	[-0.338]
Census year dummies	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Firm controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Location controls	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Distance to schools	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
R-squared	0.300	0.186	0.271	0.331	0.210	0.291	0.402
Observations	645,137	611,748	33,389	457,768	184,637	537,630	107,507
F-statistic (first-stage)	183.0	181.6	120.5	183.6	75.6	276.8	28.7

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities. Column (6) excludes observations from the two half-cantons of Basel (BS and BL), as well as from the cantons of Geneva (GE) and Ticino (TI).

(3). Consequently, $\widehat{\beta} = -1.445$ in column (1) and $\widehat{\beta} = -11.523$ in column (3) associates a ten percentage point increase in the share of CBWs with a decrease in A_i of 17.5% among firms in the baseline sample and 21.8% among firms with 50 or more employees.

Another contrast of interest is based on the observation (cf. Appendix Figure B.2) that the majority of CBWs in the German, French, and Italian language regions stem from Germany or Austria, France, and Italy, respectively. These differences in CBWs' countries of origin combined with differences across the educational systems of Germany, Austria, France, and Italy likely translate into different skill compositions of CBWs across Swiss language regions. Germany and Austria have VET systems similar to that in Switzerland, so, from a firm's perspective, hiring CBWs from these countries represents a relative close substitute for investments in VET.[33] In contrast, the Swiss vocational track differs both in size and type from the French or Italian vocational tracks. This limits the possibility to substitute VET investments by hiring CBWs for firms that rely on skills provided specifically by a Swiss-type VET system and that are located in the French or Italian language regions of Switzerland.

Columns (4) and (5) of Table 1.9 test these hypotheses by applying the baseline specification (column 1) separately to a subsample of firms located in German language (column 4) and Romance language (column 5) regions.[34] In line with the hypothesize put forth above, the estimate of $\widehat{\beta} = -1.351$ among firms in the German language region exceeds the estimate of $\widehat{\beta} = -0.883$ among firms in the Romance language regions by roughly 50%. Note, however, that the confidence interval of the German-subsample coefficient is relatively large, and it is questionable whether the differences between the

[33]According to Field *et al.* (2010) a dual VET system, in which pupils spend most of their time within a firm, is especially common in the German-speaking countries Switzerland, Germany, and Austria. In contrast, France and Italy have lower overall numbers of VET pupils at the upper secondary level and low shares of combined school and work-based VET programs.

[34]The Romance language parts of Switzerland consist of the French, Italian, and Romansh language regions of Switzerland.

language regions are indeed statistically significant.

Finally, in the last two columns of Table 1.9, we check that our results are not entirely driven by those regions with a disproportionate share of CBWs (i.e. the two half-cantons of Basel and the cantons of Ticino and Geneva). Column (6) first shows the 2SLS estimate when these regions are excluded from the sample, yielding an estimate of $\widehat{\beta} = -1.412$, which is again very close to the baseline estimate, though less precisely estimated. Finally, we estimate the same specification using only firms from these regions, again yielding a similar estimate of $\widehat{\beta} = -1.342$ (note that this estimate remains statistically significant even though it uses a much smaller sample than the baseline specification).

1.6 Conclusions

In this chapter, we study the potential substitution of employers between hiring CBWs and training apprentices of firms operating in the Swiss labor market. To account for potential biases and reverse-causality issues related to employers' hiring decisions, we focus on CBWs and apply IV approach, using a firm's distance from the national border as instrument for its share of CBWs. Within our sample period, we argue that firms' location choice is largely unrelated to the distance to the border, and we thus argue that our instrument induces quasi-random variation in firms' shares of CBWs. We apply this empirical strategy to three waves of the Business Census (1995, 2005, and 2008), including a total of 645,137 observations stemming from 342,323 firms.

Our preferred 2SLS estimate implies that the increase of 1.14 percentage points in firms' shares of CBWs within our sample period from 1995 to 2008 led to a substitution of roughly 3,500 apprenticeship positions (about 2% of all apprentice positions in relative terms). This effect is statistically significant. Moreover, various robustness specifications show that this result is not sensitive to the exact specification of either the dependent or the exogenous variables, that it holds in different subsamples and for different sets of control variables, and that 2SLS estimates using differenced data also suggest our result to be robust to unobservable firm characteristics.

The present chapter is, to the best of our knowledge, the first to analyze the effect of immigration on firms' provision of apprenticeship training. The potential substitution of firms between training apprentices and hiring CBWs is not only interesting from an economic point of view but also bears political relevance, because the apprenticeship system is a key pillar within the country's educational system. Importantly, however, our results do not imply that young natives do not receive any education or training because firms substitute for apprenticeship positions. In fact, our results are consistent with the finding from other studies that immigration pushes natives into other educational tracks and occupations (e.g. Foged and Peri, 2016; Peri and Sparber, 2009). Our results do, however, suggest that immigration undermines the incentives of firms to participate in the firm-based system of education and training.

Moreover, our results also raise the more general question of substitution between firms training their own workforce versus hiring immigrant workers, whether cross-border or not. However, we believe that the answer to that question cannot directly be inferred from our results. On the one hand, we suspect that the general immigrant population in the Swiss labor market differs from CBWs in their degree of substitutability. On the other hand, resident immigrants tend to constitute a much larger share of the workforce than CBWs. This applies especially for Switzerland, where roughly 25% of the working population are foreigners, whereas CBWs make up only about 5% of the workforce. We thus motivate further analyses focusing on immigration as a whole and not limit the analysis to CBWs as this chapter does.

A Assessing instrument validity

In this appendix, we discuss the validity of our instrument, i.e. a firm's minimum distance from the national border. While we can use the first-stage estimates to assess the partial correlation between our instrument and the endogenous variable, the other two requirements for an instrument, i.e. the assumption concerning the (quasi-)random assignment of the instrument and the exclusion restriction, cannot be tested empirically (e.g. Angrist and Pischke, 2008; Wooldridge, 2010).

A.1 Endogenous location and training decisions by employers

Quite obviously, a firm's distance to the border is not random. Rather, firms decide where to locate. Within our instrumental variable approach, firms deliberately choosing their location with respect to its distance from the border may violate the assumption of the instrument being quasi-randomly assigned. If these firms differ systematically from other firms in their provision of apprenticeship positions, our 2SLS estimates will be biased. We thus first try to assess whether newly established and/or moving firms are any different in their training behavior. To that end, we estimate regressions of the following form:

$$A_i = \gamma_0 + \gamma_1 N_i + \gamma_2 M_i + \gamma_3 x_i + \gamma_4 z_{j[i]} + \psi_{t[i]} + \epsilon_i, \tag{1.4}$$

where the dependent variable A_i denotes, as in the main text, the number of apprentices trained by firm i. N_i and M_i denote whether firm i has been newly established or whether firm i has changed its location between two consecutive waves of the business census, respectively.[35] The controls are otherwise the same as those used in the regressions discussed in the main text, i.e. we include both firm- and regional-level controls.

[35]That is, the dummy variable N_i equals 1 if firm i is observed in wave t, but not in wave $t-1$. Similarily, M_i is equal to 1 if the location (i.e. municipality) of firm i in wave t differs from the same firm's location in wave $t-1$.

Table A.1: Number of apprentices and sample appearance

	Number of apprentices		
	(1)	(2)	(3)
Newly established firm (yes = 1)	-0.525***	-0.426***	-0.398***
	(0.018)	(0.024)	(0.025)
Moving firm (yes = 1)	-0.292***	-0.300***	-0.271***
	(0.020)	(0.020)	(0.015)
Constant	0.877***	0.013	1.651***
	(0.024)	(0.062)	(0.368)
Census year dummies	*No*	*Yes*	*Yes*
Firm controls	*No*	*Yes*	*Yes*
Location controls	*No*	*No*	*Yes*
R-squared	0.003	0.298	0.300
Observations	645,137	645,137	645,137

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

The resulting estimates of γ_1 and γ_2 are shown in Table A.1. Column (1) indicates that both newly established and moving firms train significantly fewer apprentices than existing and nonmoving firms. These substantial differences in the number of apprentices remain when including control variables, as evident from columns (2) and (3). Altogether, Table A.1 suggests a substantially different training behavior among newly established and moving firms. This finding threatens our identification strategy if firms' location choice also differs in these types of firms.

We thus next analyze the location choice of newly established and moving firms with respect to our instrument by running several regressions of the following form:

$$\ln(D_{j[i]}) = \delta_0 + \delta_1 N_i + \delta_2 M_i + \delta_3 x_i + \psi_{t[i]} + \epsilon_i, \tag{1.5}$$

where the dependent variable now is the log distance of a firm to the border, $\ln(D_{j[i]})$. The regressors of main interest in the above equation are again the two dummies N_i and M_i, and the controls are also the same as in Equation (1.4), except that we do not include regional-level controls here, because these are, at least in part, determined by a firm's choice of $\ln(D_{j[i]})$, i.e. the dependent variable in equation (1.5).

Table A.2: Distance to border of newly established and moving firms

	$\ln(D_{j[i]})$	
	(1)	(2)
Newly established firm (yes = 1)	-0.046***	-0.040***
	(0.013)	(0.011)
Moving firm (yes = 1)	-0.051**	-0.041*
	(0.025)	(0.024)
Constant	3.406***	3.554***
	(0.047)	(0.043)
Census year dummies	*No*	*Yes*
Firm controls	*No*	*Yes*
Location controls	*No*	*No*
R-squared	0.001	0.013
Observations	645,137	645,137

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

Table A.2 shows the resulting estimates of the two parameters of interest. Column (1), in which we do not include any control variables, reveals that newly established firms are located, in terms of driving time, about 4.6% closer to the border than existing and nonmoving firms. Similarly, moving firms are located 5.1% closer to the border than existing and nonmoving firms. In absolute terms, however, these differences are relatively small and only amount to about 1.6 and 1.8 car driving minutes. It seems questionable whether such small differences in commuting durations affect the supply of CBWs for newly established or moving firms.

However, because Table A.1 documents large differences in the number of apprentices, the fact that firms choosing their location within our sample tend to train fewer apprentices nonetheless deserves greater attention. We address such concerns in three steps. First, we control for various observable firm characteristics, e.g. industry dummies to account for industrial clusters that may have evolved dependent on the distance from the border (cf. Section 1.3.1). Second, Table A.1 shows that firms' numbers of apprentices is especially low at the beginning of their lifecycle. This motivates us to consider firms' age as an additional firm-specific control variable (as discussed in Section 1.3.1). Third, we estimate our baseline specification within various subsamples based on firms' approximate age in Section 1.5.2 in the main text.

A.2 Other threats to our identification strategy

Besides firm behavior, other channels potentially pose a threat to our identification strategy by violating the exclusion restriction. For instance, municipalities close to the border may systematically differ from regions further away. Suppose, for instance, that urbanization is higher in border regions. Presumably, this might affect the supply of apprentices because general education facilities (e.g. high schools and universities) are mostly located in or close to cities; the instrument $\ln(D_{j[i]})$ might therefore be correlated with the error term ϵ_i. To mitigate these concerns, we control for various municipality characteristics, such as size and population density. Moreover, the distance between firm j's municipality and the nearest municipality hosting any of the three educational facilities at the upper secondary level (high school, dual VET school, full-time VET school) accounts for exogenous apprentice supply shifts. All these location specific controls are described in Section 1.3.2 in the main text (descriptives are shown in Table B.1).

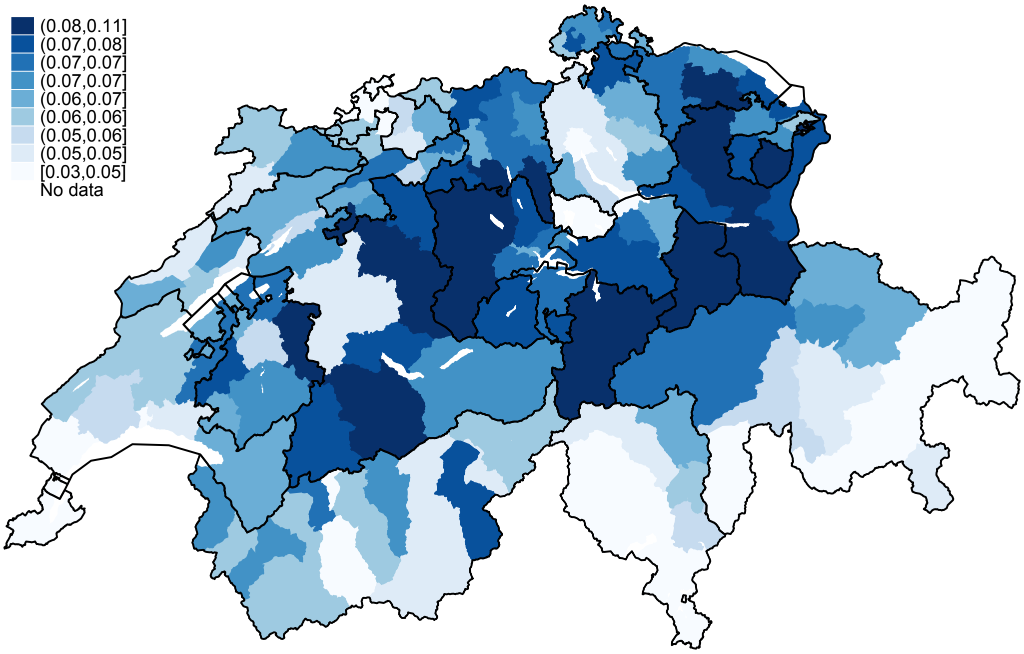
Another factor potentially interfering with the distance from the border are the number of resident immigrant workers. On the one hand, immigrants from neighboring countries might choose to settle close to the border and thus to their countries of origin. On the other hand, the labor supply of CBWs might crowd out resident immigrant workers and encourage them to move to more central regions of Switzerland. Assuming complementarity between resident immigrant workers and CBWs, both these associations would bias our results in a violation of the exclusion restriction. To mitigate concerns to that effect, we add firms' share of resident immigrant workers to the firm-specific controls (for details see Section 1.3.1).

A final threat to the exclusion restriction that we consider stems from cross-border apprentices. Any positive correlation between the supply of cross-border apprentices and the distance from the border would weaken the reduced-form association between firms' apprenticeship positions and their distance from the border (shown in Table B.4) and would bias the 2SLS estimate downwards. However, data from the Federal Statistical Office indicates that foreign apprentices commuting into Switzerland are extremely rare, almost nonexistent.[36]

[36]In the year 2012 (2013; 2014) only 37 (50; 42) out of 311,070 (313,853; 320,883) apprentices commuted from outside Switzerland to a firm within Switzerland (LVA data 2016).

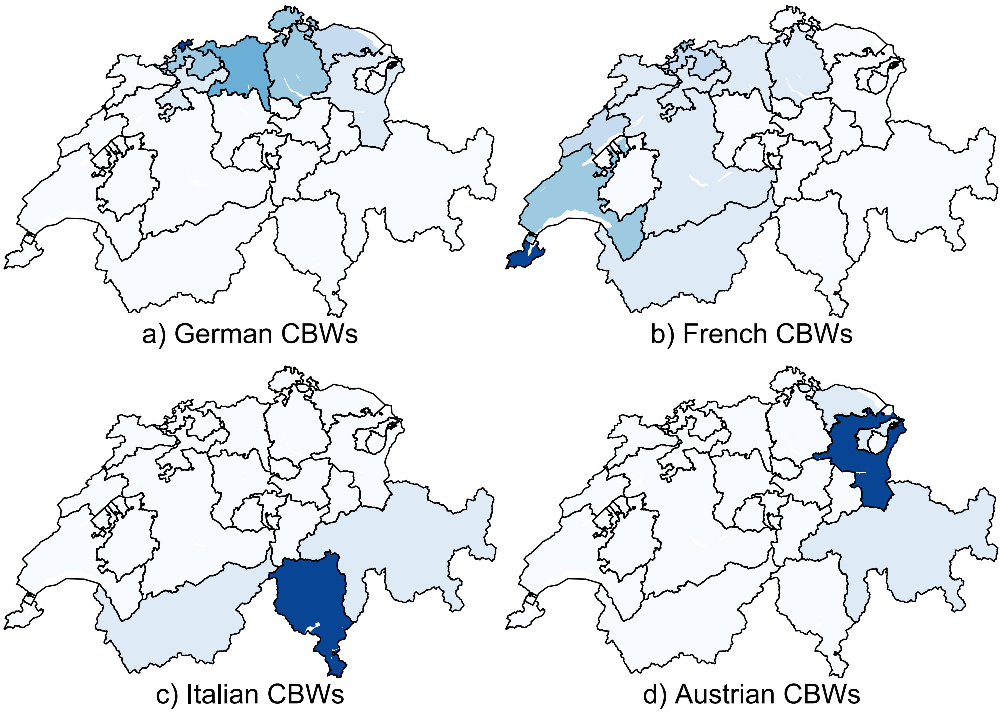
B Additional tables and figures

Figure B.1: Share of apprentices at district level



Notes: The map shows the spatial distribution of the ratio of apprentices to total work force at district level (there are 148 unique districts).

Figure B.2: Cantonal numbers of CBW by country of residence



Notes: The figure shows cantonal-level means of the absolute number of CBWs from 2002 to 2008, by their country of residency.

Source: Official Statistic on Crossborder Commuters by the Federal Statistical Office.

Table B.1: Educational attainment by population groups

	Swiss	Immigrants	CBWs
University degree	11.80	12.40	13.65
Tertiary-B degree, teaching diploma	10.97	5.12	8.79
High school degree	2.21	1.74	3.17
Vocational qualification (basic level)	63.45	41.12	53.11
Compulsory schooling	9.05	29.49	12.37
Remaining category	2.53	10.14	8.90
Total	100.00	100.00	100.00

Notes: Immigrants refers to resident immigrant workers (i.e. non-natives working and living in Switzerland). There are slight discrepancies between these figures and those derived from the data used in the analysis below (due, for example, to different sample definitions).
Source: Swiss Earnings Structure Survey (2004).

Table B.2: Descriptive statistics of individuals

	Full population[a]			Sample[b]		
Wave	1995	2005	2008	1995	2005	2008
Number of employees	3,556,506	3,713,028	4,013,453	3,328,516	3,486,617	3,780,068
Numer of apprentices	141,981	173,971	192,085	139,746	171,265	189,168
Number of CBWs	140,644	170,785	211,663	138,254	167,499	207,317
Number of immigrant workers	857,899	895,797	1,033,993	830,670	861,083	993,737
Share of apprentice	0.040	0.047	0.048	0.042	0.049	0.050
Share of CBWs	0.040	0.046	0.053	0.042	0.048	0.055
Share of immigrant workers	0.241	0.241	0.258	0.250	0.247	0.263
Share public sector	0.183	0.166	0.157	0.190	0.173	0.163
Share third sector	0.694	0.735	0.736	0.686	0.730	0.731
Number of firms	373,472	375,392	388,324	212,081	213,106	219,950

Notes: Immigrant workers refers to the number (share) of resident immigrant workers (i.e. non-natives living and working in Switzerland). [a]All firms operating in the second or third sector. [b]All firms operating in the second or third sector with three or more employees.

Table B.3: First-stage estimates

	Share of CBWs				
	(1)	(2)	(3)	(4)	(5)
$\ln(D_{j[i]})$	-0.067***	-0.067***	-0.068***	-0.054***	-0.056***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)
Census year dummies	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Firm controls	*No*	*No*	*Yes*	*Yes*	*Yes*
Location controls	*No*	*No*	*No*	*Yes*	*Yes*
Distance to schools	*No*	*No*	*No*	*No*	*Yes*
R-squared	0.132	0.134	0.159	0.242	0.244
Observations	645,137	645,137	645,137	645,137	645,137
F-statistic (first-stage)	201.6	201.8	174.4	170.7	183.0

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

Table B.4: Reduced-form estimates

	Number of apprentices				
	(1)	(2)	(3)	(4)	(5)
$\ln(D_{j[i]})$	0.118***	0.119***	0.133***	0.072***	0.080***
	(0.032)	(0.032)	(0.026)	(0.017)	(0.016)
Census year dummies	*No*	*Yes*	*Yes*	*Yes*	*Yes*
Firm controls	*No*	*No*	*Yes*	*Yes*	*Yes*
Location controls	*No*	*No*	*No*	*Yes*	*Yes*
Distance to schools	*No*	*No*	*No*	*No*	*Yes*
R-squared	0.000	0.001	0.298	0.300	0.300
Observations	645,137	645,137	645,137	645,137	645,137

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipalities.

Chapter 2

Train to meet the norm: Local norms towards private engagement and the provision of apprenticeship positions

joint work with **Andreas Kuhn and Jürg Schweri**

2.1 Introduction

> *Marge, I can't wear a pink shirt to work. Everybody wears white shirts. I'm not popular enough to be different...*
>
> Homer Simpson, *Season 3, Episode 1*

Vocational education and training (VET) is Switzerland's most important upper-secondary track, with over 60% enrolments per cohort (FSO, 2017). This VET system relies heavily on voluntarily participating firms which hire apprentices, pay their wages, and instruct and teach them at the workplace. Gehret *et al.* (2019) estimate the total VET costs borne by firms to be up to five billion Swiss Francs in 2016. Even though firms also profit financially from their apprentices being partly productive, many firms incur substantial net costs or realize only small net benefits during an apprenticeship. Classical profit-maximizing behavior seems therefore unable to entirely explain firms' motivation to offer apprenticeship positions repeatedly every year.

In this chapter, we therefore present a complementary explanation for firms' provision of apprenticeship positions based on prevalent norms. More specifically, we hypothesize that a locally persistent norm favoring private over state provision of goods enhances private investments in these goods, in our case VET. This claim stems from the study by Kuhn *et al.* (2019), who show that firms' decision to participate in the apprenticeship system is influenced by local attitudes towards the role of the state. In the context of firms' VET investments in Switzerland, local norms might thus help to explain why some firms train and others do not, even though they have similar characteristics and would thus bear similar monetary costs from offering an apprenticeship.

In contrast to Kuhn *et al.* (2019), we use administrative data and a different empirical approach. Empirically, we follow the approach proposed by Eugster *et al.* (2017) and profit from a language-cultural border within Switzerland evoked by its multilingualism. Because norms, as a part of culture, are mainly transmitted by language (Guiso *et al.*, 2006), they likely spread relatively homogeneously throughout language regions but deviate sharply at language borders. Applying a fuzzy spatial regression discontinuity design at this border thus permits us to evaluate the association between diverging norms concerning private provision of goods and firms' training incidence. Crucial for the validity of this approach, Switzerland's language borders mostly do not coincide with institutional borders, which limits bias emerging from confounding factors.

Three main findings emerge from our empirical analysis. First, voting results yield remarkable differences in attitudes to private engagement in the provision of goods among Swiss language regions. German speakers regularly oppose state interventions and value the private provision of goods, whereas Romance-speaking (French, Italian, and Romansh) voters favor state engagement in the provision of the same goods. This contrast appears both as a difference in average voting results across language regions and as a discontinuity at the language border. Second, the difference in the share of training firms on either side of the language border is remarkable. As many as 31.9% of all firms across German-speaking regions train apprentices, but only 25.7% of all firms in Romance-speaking regions train. This difference decreases slightly when evaluated close to the language border but remains statistically significant. Third, our preferred estimate associates a one-standard-deviation difference in the norm demanding private provision of goods with a 3.6 percentage point higher share of training firms. These results are robust throughout various specifications and subsample estimations. Moreover, the estimates remain widely unaltered when controlling for canton fixed effects, firm and location characteristics, and demand for apprenticeships. Altogether, we argue that norm-guided behavior is a complementary explanation for why some firms train apprentices and others do not. This finding adds to the existing literature on firms' training behavior, which is based mostly on financial considerations and is discussed in more detail in Section 2.2.1.

The rest of the chapter is organized as follows. Section 2.2 provides some background information on the Swiss apprenticeship system and on the country's different language-cultural regions and introduces the theoretical framework on which we base our norm hypothesis. Section 2.3 presents the main data sources and discusses the construction of the key variables, and Section 2.4 sets out our empirical strategy. Section 2.5 discusses our main results, including several robustness checks, and Section 2.6 presents additional evidence on the underlying mechanisms. Section 2.7 concludes.

2.2 Background

2.2.1 The Swiss apprenticeship system

Over 60% of a Swiss youth cohort attend VET, which makes this the most popular track at the upper-secondary level (FSO, 2017). Usually, youngsters apply immediately after compulsory schooling, at around the age of 16, for an apprenticeship position offered voluntarily by a firm. If successful, the apprentice works and trains three or four days per week within the training firm and spends the other one or two days at a vocational school. Most apprenticeships last for three years, some for two or four. While vocational schools are publicly financed, participating firms cover the costs incurred during the apprentices' period of training at the firm. These costs amount to roughly five billion Swiss Francs and consist primarily of apprentice wages, time spent on instruction, and training equipment.[1] Participating firms also profit from their apprentices, who work partly productively during the training period. Roughly one-third of the training firms incur net costs over the whole training period, whereas about two thirds of all training firms in Switzerland earn net benefits. However, these benefits are often small and unpredictable at the beginning of the training period for many of the firms ever (Gehret *et al.*, 2019).

Beside financial gains during an apprenticeship, training firms can try to recoup their training investments by retaining fully trained apprentices within their firm after the training period. This allows them to save search and recruitment costs (e.g. Blatter *et al.*, 2016; Gehret *et al.*, 2019; Wolter and Strupler, 2012).[2] However, standardized curricula in Switzerland ensure up-to-date apprenticeship training content in firms and the transferability of accumulated skills to other firms. As a likely consequence of this policy, Müller and Schweri (2015) find no wage penalties for former apprentices leaving their training firm within one year after completing VET but remaining within the same occupation. They thus conclude that human capital accrued during an apprenticeship in Switzerland is widely transferable

[1]Public costs for VET amount to roughly three billion Swiss Francs and primarily include costs for VET schools and costs for final exams, https://www.bfs.admin.ch/bfs/de/home/statistiken/bildung-wissenschaft/bildungsindikatoren/bildungssystem-schweiz/themen/investitionen-und-kosten/ausgaben-berufsbildung.html.

[2]Moreover, Acemoglu and Pischke (1999) model how investments in general skills might be beneficial for firms if labor market frictions allow them to employ trained workers at wages below their marginal productivity. Based on this work, Mühlemann *et al.* (2010) show how the differing strengths of labor market regulations in Switzerland and Germany can partly explain observed differences in firms' VET investments across the two countries.

within the occupational field.[3] This, combined with Switzerland's weakly regulated labor market, makes it relatively easy for firms to poach fully trained apprentices from other firms and thus limits firms' possibilities to recoup training investments after the training period. Accordingly, Mühlemann and Wolter (2011) document that the labor market density measured as local number of firms per hectare in the same industry lowers firms' training probability by 0.18 percentage points.

Altogether, we surmise that financial incentives alone, whether during the training period or after it, cannot entirely explain the high and constant apprenticeship provision by Swiss firms. In Section 2.2.3, we will thus put forth a complementary explanation for this pattern based on locally prevalent norms. Before doing so, however, we introduce Switzerland's language regions in the next, Section 2.2.2, and demonstrate why these are well suited for studying the association between diverging norms regarding the private provision of goods and firms' apprenticeship provision.

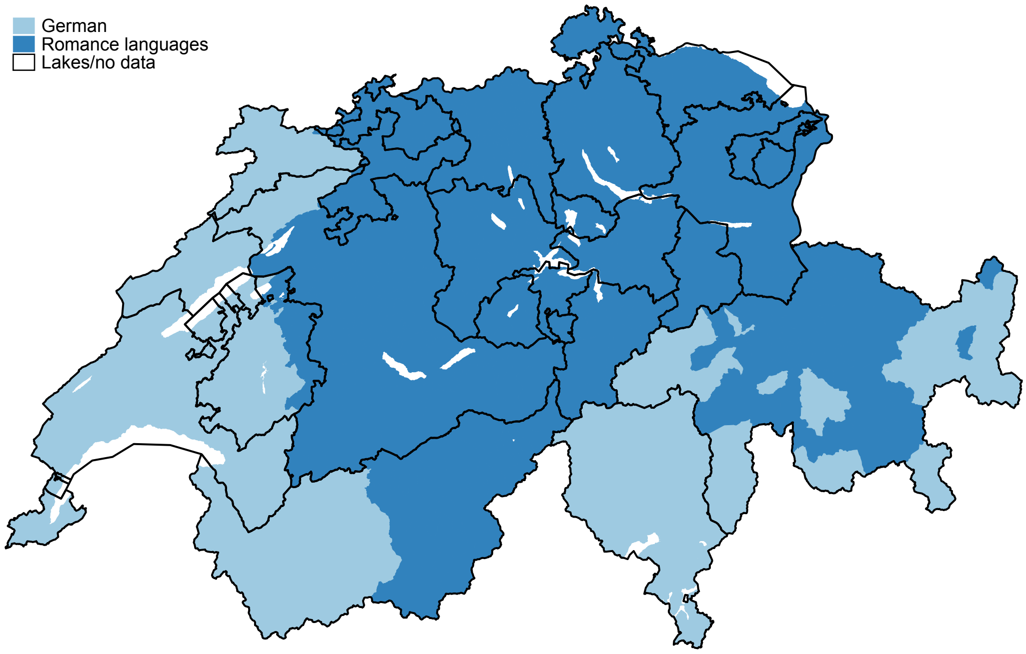
2.2.2 Language-cultural regions

Switzerland is a multilingual country with four official languages. According to the 2000 Population Census, German is the first language for a majority of its resident population (63.7%). In the western part of Switzerland, bordering France, French is the official language and at 20.4% the second most prevalent first language in Switzerland overall. The Italian language (6.5%) is closely associated with the Canton of Ticino in the south of Switzerland, where it is the sole official language. The fourth official language, Romansh (0.5%), is only spoken in parts of the Canton of Grisons. This canton in southeastern Switzerland is the only trilingual canton; in addition to the Romansh population, a majority speaks German and another minority speaks Italian. The remaining 8.9% of residents state that they do not speak any of the four official languages.

For our empirical analysis, we group these different language areas into

[3]However, workers who change occupation face a wage differential of roughly 9% (Müller and Schweri, 2015).

Figure 2.1: Romance- and German language regions of Switzerland



Notes: This is a map of Switzerland showing the breakdown into the two main language-cultural regions at the level of municipalities (see also Appendix Figure B.1, which shows the four language regions). A municipality is considered a German- or Romance- speaking municipality if the majority of its inhabitants speak German or one of the Romance languages, respectively, according to the Population Census from the year 2000. The black lines delineate cantonal borders.

two main language regions, one German speaking and the other Romance speaking (i.e. French-speaking region, Italian-speaking region, and Romansh-speaking region).[4] Figure 2.1 shows the two language regions along with the cantonal borders (see also Appendix Figure B.1). Cantons are Switzerland's most important institutional entities, enjoying a considerable amount of autonomy, for instance over taxation and the educational system. Crucially

[4]A municipality is considered German-speaking or Romance-speaking when the majority of its inhabitants speaks German or a Romance-language respectively. According to the 2000 census, 71.7% of the resident population lived in German-speaking municipalities, 23.6% in French-speaking municipalities, 4.4% in Italian-speaking municipalities, and 0.3% in Romansh-speaking municipalities. Brügger *et al.* (2009), Eugster *et al.* (2017), and Cottier (2018), which papers are discussed shortly in Section 2.2.3, all proceed similarly.

for our econometric approach, the language border runs largely through the multilingual Cantons of Bern, Fribourg, Grisons, and Valais, and mostly does not coincide with cantonal borders, which implies that it is possible to net out institutional differences by including fixed effects at the cantonal level. Moreover, neither important geographical territories nor economic hubs spread homogeneously across one or the other language regions. The dominant language border (German/French) runs from north to south in western Switzerland, while the most important geographical separators run from west to east: the Alps (south) and Jura (north) mountain ranges. In between them, the densely populated Swiss Plateau ranges from the far-west French-speaking part to the far-east German-speaking part and hosts the majority of Switzerland's people and economic activities.

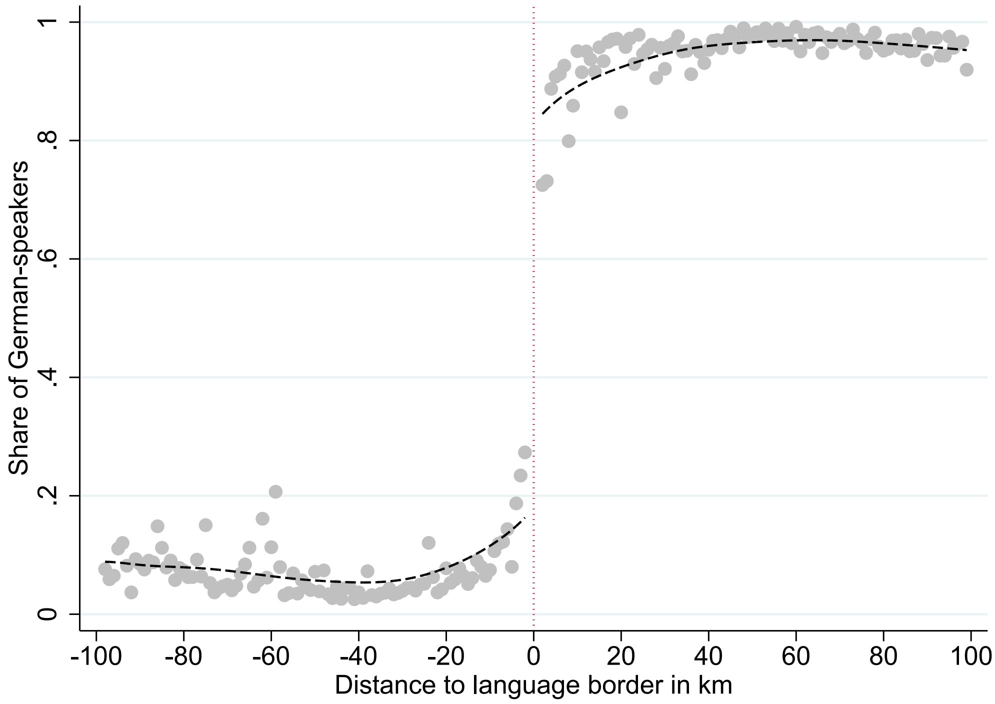
Figure 2.2 plots the share of German speakers with respect to the distance from the language border and displays a clear discontinuity around the language border.

2.2.3 Local norms favoring private over public engagement

Social scientists, including economists, have argued since the establishment of their field that human behavior is, in addition to rationality, based on beliefs, customs, and norms. The early works of Adam Smith and John Stuart Mill bear witness to this (Mill, 1929; Smith, 1761).[5] More recently, Fehr and Fischbacher (2004) state that "cooperation in human societies is mainly based on social norms" (p.185). According to them, this is even true in modern societies in which cooperation is often enforced by legal rules, because rules lack their normative legitimacy if social norms do not back them. Generally, social norms are prevalent when individuals conform their behavior to the standard behavior within a group, e.g. a family, a group of peers, or a whole society (e.g. Burke and Young, 2011; Ellickson, 1999; Elster, 1989b). Very similarly, Postlewaite (2011) use the term *social norms* to describe why agents

[5] Guiso *et al.* (2006) for an introduction to research on the effect of culture on economic outcomes and a historical overview of it.

Figure 2.2: Discontinuity in the share of German-speakers at the language border



Notes: The figure shows the share of German speakers among the resident population, aggregated in bins of 1km width, by their distance from the language border in terms of actual travelling distance. The dashed line shows smoothed values from a locally weighted regression.

in one group behave differently from agents in another, even when the groups are similar in their composition and the physical environment they face (for empirical applications see Cole *et al.*, 1995, 2001). According to Burke and Young (2011), social norms emerge and are sustained through a positive and thus self-enforcing *feedback loop* echoing between individuals' behavior and the reactions they receive to it from their peers. Because it is not ex ante clear which norms materialize, this feedback loop leads to *local uniformity* within groups but can evolve into substantial behavioral differences across groups (*global diversity*) (for an empirical application see Burke and Heiland, 2007). In the view of Elster (1989a) social norm do not necessarily require feedback to sustain. However, he agrees that norms, to be social, must be shared by others and partly sustained by their approval or disapproval.[6]

Altogether, the key feature of social norms seems to be their materialization through social interactions within social groups. Based on the fact that social interactions, as with any other cultural expression, most often occur in one form of language or another (Guiso *et al.*, 2006), we argue that Switzerland's language regions also form separate regions in terms of culture and consequently in terms of norms.[7]

Other papers embedded in the context of the different language regions in Switzerland argue similarly but focus on other cultural and economic differences. Eugster and Parchet (2013) document deviating preferences for tax competition at the German-French language border. Both Cottier (2018) and Eugster *et al.* (2017) find differences in work attitudes at the Swiss language border that translate into contrasting levels of unemployment durations and diverging retirement decisions. Finally, Steinhauer (2018) shows that different norms regarding working mothers along Switzerland's German-French language border correspond to varying work participation rates and fertility among women.

In this chapter, we focus on a norm preferring private engagement in the provision of goods in contrast to state provision. As we will show, this norm is more prevalent among German-speaking regions than among Romance-speaking regions. We argue that in the Swiss institutional setting, where private (firm-provided apprenticeships) and state (high schools, school-based apprenticeships) investments in education are to some extent substitutes, the spatially varying dispersion of this norm eventually translates into contrasting levels of firm-provided apprenticeships. Specifically, we hypothesize the probability of firms' providing apprenticeships is higher among regions where people favor private over state provision of goods.

[6]He adds that social norms can also be sustained by personal feelings, such as anger, guilt, and embarrassment.

[7]In this context, one should note that the aim of this chapter is not to evaluate any direct association between language and firms' apprenticeship provision; we genuinely believe there is none.

2.3 Data

2.3.1 Firm-level data from the Swiss Business Census

Our main data source is the Swiss Business Census, which covers all Swiss firms operating in the second or third sector, and we use the four most recent waves (1998, 2001, 2005, and 2008) for our empirical analysis.[8] There are two reasons to use several waves of the Business Census. First, our empirical approach is very observation intensive (cf. Section 2.4). Second, using several waves also ensures that the time period from which we draw firm-level observations approximately overlaps with the time frame implicit in our measure of private engagement, discussed in Section 2.3.2 below. The only sample restriction that we impose is that we exclude very small firms with less than three employees, because only a tiny fraction of these firms train any apprentices. For that reason, these firms do not add much useful variation in the dependent variable. Therefore, as previous studies on similar subjects (e.g. Gehret *et al.*, 2019; Kuhn *et al.*, 2019; Wolter and Strupler, 2012) have done, we decided to remove these firms from the sample; however, we will show that our results are robust to this sample restriction (cf. Section 2.5.3). This restriction narrows our sample down to 842,146 observations at the firm$\times$census-year level.

Our main dependent variable is a dummy variable T_{it}, indicating whether firm i trains at least one apprentice in census year t.[9] In the full sample, about 30% of the firms are involved in apprenticeship training, as shown in panel (a) of Table 2.1. The table also shows descriptives for a subsample of firms located within 20 kilometers from the language border, henceforth referred to as the regression discontinuity sample or RD sample. As we

[8]The Business Census was superseded by an alternative firm-related data collection (STATENT), which, however, does not include information on apprentices anymore.

[9]In the empirical analysis, we essentially focus on the cross-sectional variation in T, mainly because the main regressor does not vary across time (cf. Section 2.3.2). We thus suppress the t index in what follows, but we include fixed effects for census years in most of the regressions reported below.

Table 2.1: Descriptive statistics

	Full sample		RD sample			
	Mean	SD	Mean	SD	Aggregation	Data source
(a) Main variables						
Training firm (yes = 1), T_i	0.301	0.459	0.322	0.467	Firm	Business Census
Private engagement, N_j	0.600	0.096	0.568	0.086	Municipality	FSO, Votes[a]
(b) Firm-level characteristics						
Number of employees	16.3	65.7	13.7	38.0	Firm	Business Census
% For-profit	0.84	0.37	0.82	0.38	Firm	Business Census
% Public	0.11	0.31	0.13	0.33	Firm	Business Census
(c) Location characteristics						
Log(number of firms)	8.03	0.74	7.60	0.63	LM-region[b]	Business Census
Log(same-sector firms)	6.22	0.71	5.85	0.63	LM-region	Business Census
Log(Inhabitants)	7.23	1.21	7.05	1.09	Municipality	FSO[c]
Population density	366.9	671.7	269.6	477.1	Municipality	FSO
% Employed	0.61	0.02	0.61	0.01	LM-region	FSO
% Employed in 2nd sector	0.34	0.17	0.34	0.17	Municipality	Business Census
% Employed in 3rd sector	0.49	0.18	0.47	0.17	Municipality	Business Census
Median income in CHF	6899.0	593.1	6748.9	456.2	LM-region	SLFS[d]
(d) Apprenticeship demand controls						
Distance to high school	12.9	9.5	17.0	11.2	Municipality	FSO [e]
Distance to dual VET school	9.8	7.4	12.2	9.5	Municipality	FSO
Distance to school-based VET	14.3	9.7	15.7	11.5	Municipality	FSO
% Age 15-25	0.12	0.03	0.11	0.04	LM-region	Pop. Census
Observations (firm×census-year)	842,146		69,619			
Observations (municipalities)[f]	2,315		371			

Notes: [a]Federal Statistical Office (FSO), Statistics of elections and votes. [b]106 Labor market regions defined by the FSO. [c]FSO, Portraits of the communes. [d]Swiss Labor Force Surveys 2010-2014. [e]FSO, Educational institutions, via email. [f]For this table we only consider municipalities with at least one firm observation.

discuss in Section 2.4 below, most of our empirical analysis focuses on this much smaller subsample of 69,619 observations. Within this subsample, the share of training firms is 32.3%, somewhat higher than in the overall sample (see also Appendix Table B.1, which breaks down the descriptives for the RD sample by language region).

2.3.2 Measuring local norms towards private engagement

To measure the local norm favoring private over public engagement, we use municipal voting results from eight national-level plebiscites that took place in Switzerland between 1986 and 2014 (Appendix Table B.2 contains some additional information regarding these votes). Although dealing with quite distinct substantive issues, these eight votes all share the overarching question of whether citizens should confer more responsibility on the state. For example, one of these votes (vote 503; cf. Table B.2) called for state action in the apprenticeship system (in that period, the Swiss economy was in recession, and firms' supply of apprenticeship positions temporarily could not meet demand).

In what follows, $N_{j[i]}$ will refer to our measure of the local norm favoring private engagement, measured by the mean share of votes in support of more private engagement, and thus less engagement by public authorities, in municipality j within which firm i is located. Note that there is no temporal variation in this measure, because we aggregate the votes from the different years. For each vote, we use the share of the votes in support of more private engagement. Because plebiscites tend to demand additional state action, we thus use the share of no-votes in the construction of our measure in seven out of the eight plebiscites (as detailed in Appendix Table B.2). As shown in panel (a) of Table 2.1, the mean share of votes in support of more private engagement equals about 60% in the full sample; focusing on the RD sample, the mean value of N_j is 56.9%. This implies that there is a majority in support of private, rather than public, engagement in both the full and the RD sample.

Reverse causality in the norm measurement

The strategy described above for measuring the local norm $N_{j[i]}$ raises one particular concern: the norm favoring private engagement for the provision of a good prevalent in municipality j, $N_{j[i]}$, might be endogenous to the actual level of this good in municipality j. For instance, people living in regions where firms provide apprenticeship positions on a large scale may not recognize any need to strengthen VET and therefore disapprove of state investments in VET. We attenuate concerns about this potential reverse causality issue in two steps.

First, we decide not to focus exclusively on votes dealing explicitly with VET policies (votes nr. 340 and 503) to measure $N_{j[i]}$.[10] Since the approval rates across these other votes are likely independent of the apprenticeship provision of local firms, using non-VET votes helps to measure $N_{j[i]}$ in a way that it is exogenous to the local level of apprenticeship provision of firms.[11]

Second, we analyse data from an exit poll related to the "apprenticeship initiative" (vote 503, cf. Appendix Table B.2), which explicitly demanded that the state should take more responsibility within the apprenticeship system in Table 2.2. Apparently, the share of voters who at least partly approved of the apprenticeship initiative due to a perceived lack of apprenticeship positions was roughly equally prevalent across language regions, though actual apprenticeship provision by firms is, as we show later, lower in Romance-speaking regions. This suggests that lower levels of firms' apprenticeship provision in Romance-language regions did not trigger the higher approval rate in the apprenticeship initiative among Romance-speakers than among German-speakers.[12]

[10]The main result of our analysis (cf. Section 2.5.3 below) are not sensitive to whether $N_{j[i]}$ is measured with the two votes dealing with VET policies or with the other six votes.

[11]Note that a spatial correlation in the provision of VET and the goods discussed in the other votes, e.g. because the general financial endowment across regions varies, appears unlikely because the other votes demanded public engagement in the provision of goods which were, at the time the votes took place, more or less evenly accessible throughout all Switzerland. Mainly this was the case because they were already provided by the national state, though at generally lower levels than the votes demanded it.

[12]In contrast, and in line with the arguments in Section 2.3.2, Table 2.8 suggests that

Table 2.2: "Apprentice initiative", reason for 'Yes': more apprenticeship positions

Mean answer				Conditional[a]
All	Romance	German	Difference	difference
0.624	0.614	0.630	0.016	0.060
(0.034)	(0.054)	(0.044)	(0.070)	(0.079)
[202]	[83]	[119]		

Notes: This table shows the share of yes-voters indicating they approved vote 503 either because "Pupils need apprenticeship positions" or "There are not enough apprenticeship positions" or "It is a general right to have an apprenticeship position" or "More possibilities for apprentices". The square brackets indicate the number of observations. [a]OLS-regression of a dummy with 1 for yes-voters stating one of the motives above on a German-speaking region dummy, demographic (age, age squared, sex, civil status, employment status) and municipality (22 categories) characteristics.
Source: VOX exit poll.

2.3.3 Distance from the language border

As we will detail below (cf. Section 2.4), our empirical approach focuses on firms that are located relatively close to the language border. For that reason, it is also crucial that we are able to precisely determine firms' proximity to the language border. To do so, we use additional data from the Federal Office for Spatial Development (which uses the data for traffic planning on behalf of the Federal Government). These data record actual average travelling distances by car (i.e. not simply linear distances) from a given municipality to any other municipality as of the year 2010 and based on actual commuting patterns.

These additional data on travelling distances can be merged with the firm-level data from the Business Census at the municipality level. To compute a

a more favorable attitude towards state engagement in the provision of VET contributed to the higher share of yes-votes in Romance-language regions.

given firm's distance from the language border, we proceed as follows. First, based on the Population Census from the year 2000, we define all of the 2,352 Swiss municipalities with a majority of Romance (German)speakers as belonging to the Romance- (German-) language region.[13] Then, for every Romance-speaking municipality, we keep the shortest travelling distance to any German-speaking municipality situated at the language border and vice versa, and we set the distance to negative values for all Romance-language municipalities. Across all Swiss municipalities, the travelling distance from the language border for Romance-speaking municipalities ranges from -1.4 to -141.8 kilometers and from 1.4 to 160.8 kilometers for German-speaking municipalities (see Appendix Figure B.2 for a graphical representation of this variable).

2.3.4 Control variables

The remainder of Table 2.1 contains descriptives for most of the covariates used later on (except for a firm's industrial affiliation, to save space). As indicated in the last column of the table, all firm-level characteristics are taken from the Business Census, and the location-level characteristics and variables controlling for the demand for apprenticeship positions are for the most part derived from various additional sources of data.

Firm-level characteristics

Firm-level characteristics available in the Business Census include a firm's number of employees, its industrial affiliation, and its legal status. In the empirical analysis below, we include a firm's number of employees and its square, 19 dummies representing the industry of a firm, and three dummies mapping the legal status of a firm as control variables.

[13]As Eugster *et al.* (2017) did, we exclude the two bilingual cities of Biel/Bienne and Freiburg/Fribourg because these cannot be assigned unambiguously to a language-cultural region.

Locational characteristics

Because of the spatial nature of our empirical design, we also include several location-level controls at different levels of regional aggregation (cf. Table 2.1). To map the general structure of a labor market, we include the log number of residents and the population density at the municipality level, the share of the resident population in employment, the share of people working in either the second or the third sector, and the median income at the level of local labor markets. We also control for the total number of firms and the number of firms within the same industry in the respective labor market. These last two locational characteristics map a labor market's density and account for three potential mechanisms affecting a firm's willingness to train. First, given the number of young people, other locally present firms potentially lower the number of apprenticeship applicants for a specific firm. Second, instead of training and retaining apprentices, a firm can meet its demand for skilled workers by poaching trained apprentices from other same-sector firms (e.g. Mühlemann and Wolter, 2011). And third, a firm might lower its training investments because it fears such poaching by local competitors.

Demand for apprenticeship positions

One concern with our identification strategy is that there may also exist a discontinuity in the demand for apprenticeship positions at the language border. To study the association between local norms and firms' provision of apprenticeship positions, we would like to observe a firm's training intention no matter whether it ultimately employs an apprentice or not. In contrast, our dependent variable, a firm's training status, captures the outcome on the apprenticeship market. We target this issue mainly by also including variables which control for variation in the demand for apprenticeship training.[14] Specifically, we include a firm's minimum distance from any of the

[14] We are not able to fully control for differences in the demand for apprenticeship positions, but at least we can check whether the parameter estimates change when we also include demand-side controls. Additional checks related to differences in the demand for

three main upper-secondary schooling types (i.e. vocational school, full-time vocational school, and high school) in our model. Moreover, we also control for a firm's potential number of apprenticeship applicants by including the proportion of people between the age of 15 and 25 in the labor market where the firm operates.

2.4 Empirical design

The sharpness in the language border that runs across the country opens up the possibility of estimating the association between local norms favoring private engagement and firms' training behavior using a spatial regression discontinuity design. This approach was first used by Brügger *et al.* (2009) and Eugster *et al.* (2017) to estimate the language-border contrast in attitudes towards and demand for social insurance and in unemployment duration, respectively.[15]

2.4.1 Basic setup

To motivate our own empirical approach, we start with the goal of estimating the reduced-form contrast in firms' training probability at the linguistic-cultural border in Switzerland, i.e. the language-border contrast in the probability of providing apprenticeship positions among employers. Our starting point is thus the following simple regression:

$$T_i = \alpha_0 + \alpha_1 G_{j[i]} + \epsilon_i \quad \forall \quad i \in \{i : |d_{j[i]}| \leq bw\}, \tag{2.1}$$

apprenticeship training are discussed in Section 2.6.2 below.

[15]Using a similar design, Basten and Betz (2013) estimate the effect of religion on preferences towards leisure and towards governmental intervention; Egger and Lassmann (2015) study the impact of a common native language on trade, and Cottier (2018) explores the effect of culture on labor-force exit of older workers. Examples of studies using a spatial regression discontinuity design outside the context of the Swiss language border include Keele *et al.* (2015), who look at whether ballot initiatives have an impact on voter turnout, MacDonald *et al.* (2016), studying the effect of private police on crime, and Hidano *et al.* (2015), who estimate how information about seismic hazards affects property prices in the Tokyo area.

where the binary dependent variable T_i indicates whether firm i trains any apprentices or not, and with $G_{j[i]}$ also a dummy variable denoting whether a majority of people speak German in the community j in which a given firm i is located. Note that the estimation sample only includes firms located in a community within a small band along the language border (i.e. the RD sample). Specifically, we focus on those communities not farther away, measured in travelling distance $d_{j[i]}$ (in kilometers), from the language border than a given distance bw in either direction. In the empirical results reported below, we mostly focus on the case where bw is set to 20 kilometers, implying that the language border can be reached in about 15 minutes or less by car.[16] Setting $bw = 20$ implies that our analysis focuses on those 379 communities located along the language-cultural border (about 16.1% of all communities), which provide about 8.3% of all observations (cf. columns 3 and 4 of Table 2.1). Parameter α_1 quantifies the difference in the training intensity among firms close to the language border but located on different sides of the language border and therefore exposed to different norms regarding the role of firms in the provision of training. For that reason, and assuming that individuals in the German-speaking regions are more favorable towards private engagement, we expect the training probability among firms to be lower on the Romance-speaking side of the language border, and thus we expect to find that $\alpha_1 > 0$.

2.4.2 Non-standard features due to the spatial nature of the data

Equation (2.1) relies on the assumption that observations from either side of the language border are on average identical in every respect except

[16]Remember that the selection of firms close to the language border can only be implemented indirectly by selecting municipalities close to the language border because travelling distance is only available at the community level (cf. Section 2.3). This, however, potentially biases conventional inference methods (e.g. Moulton, 1990). For that reason, we show standard errors that are clustered at the community level throughout in the regressions below.

that they belong to different cultural regions and are therefore exposed to different norms regarding the role of the state.[17] While observations close to the threshold but from different sides of the cutoff are usually assumed to be comparable in this sense in conventional applications of the regression discontinuity design (RDD), at least if individuals lack perfect control over the assignment variable, this assumption is clearly less innocuous in the context of a spatial RDD because the spatial distribution of firms and larger aggregates (e.g. cities) is nonrandom (Keele *et al.*, 2015, among others, make a similar argument).[18] In most of our regression specifications, we therefore also include control variables at both the firm- and the regional-level, denoted by X_i and $Z_{r[i]}$, respectively:

$$\begin{aligned} T_i = \; & \alpha_0 + \alpha_1 G_{j[i]} + \alpha_2 X_i + \alpha_3 Z_{r[i]} + \lambda_{t[i]} + \psi_{c[i]} + \epsilon_i \\ & \forall \;\; i \in \{i : |d_{j[i]}| \leq bw\}, \end{aligned} \tag{2.2}$$

Moreover, in most of our specifications, we also include a set of census-year dummies, denoted by $\lambda_{t[i]}$, because we pool data from several waves of the Business Census. The census-year fixed effects pick up any shifts in the mean training incidence among firms over time. More importantly, we also include a set of regional fixed effects at the cantonal level, denoted by $\psi_{c[i]}$. These allow us to control for institutional differences between the cantons (note that it is possible to include these fixed effects because there are several bilingual cantons; as evident from Figure 2.1). The inclusion of the cantonal fixed-effects may turn out to be important because cantons have considerable leeway in both educational and financial policy. Again, parameter α_1 quantifies the difference in firms' propensity to offer apprenticeship positions at the language-cultural border, but in this case conditional on observable firm- and community-level controls. As above, we still expect that $\alpha_1 > 0$.

[17] Table B.1 in the Appendix broadly confirms this.

[18] In our case, for example, larger communities tend to have a lower training intensity. The location of these communities is presumably rooted in both geography and history, but has most likely nothing to do with distance from the language border by itself. However, whatever their distance from the language border, these communities will have a significant impact on the RDD estimates. For this reason, we believe that it is important to include additional control variables.

We further deviate from the conventional setup used in an RDD in that we do not include $d_{j[i]}$ itself as a control variable, at least not in our baseline specifications. There are several partially intertwined reasons for not doing so.[19] The first and most important argument is that we do not expect any direct effect of $d_{j[i]}$ on the likelihood of providing apprenticeship positions because each community basically belongs to either the German- or the Romance-speaking cultural region, whatever their distance from the language border.[20] Secondly, and related to the first argument, it appears that any trend in the training intensity over distance to the language border is driven by variation in firm- and/or community-level characteristics rather than by the distance from the language border. Third, and finally, because we generally use quite a narrow band, including a potentially mis-specified trend in the assignment variable would probably do more harm than good (Gelman and Imbens, 2019, and also illustrated in Appendix Figure A.2).

2.4.3 Fuzzy RD estimates

Estimation of equations (2.1) and (2.2) further assumes that there is a sharp discontinuity in attitudes towards the role of the state at the language border (which, in our context, would require that the share of supporting votes jump from 0 to 100 percent). While the evidence provided in Section 2.5.1 unambiguously shows that there is a strong discontinuity in these attitudes at the language border, it also shows that the discontinuity is considerably less than 100 percent. To properly take this issue into account, we first estimate the contrast in local attitudes towards the role of the state at the language border. The corresponding first-stage regression is given by the following

[19]We show, however, that our estimates are robust to the inclusion of $d_{j[i]}$ as an additional control. These additional results are presented and discussed in Appendix A.

[20]Several pieces of evidence support this claim. First, the discontinuity in the share of Romance-speaking individuals is quite sharp at the language border, as shown in Appendix Figure 2.2. Second, Figure 2.3 shows that the mean value of $N_{j[i]}$ is relatively constant on both sides of the language border. Finally, we find that $d_{j[i]}$ has no direct effect on either $N_{j[i]}$ nor on T_i, as shown in Appendix Tables A.1 and A.2.

equation, whose right-hand side is the same as in Equation (2.2):

$$\begin{aligned} N_{j[i]} = {} & \pi_0 + \pi_1 G_{j]i]} + \pi_2 X_i + \pi_3 Z_{r[i]} + \lambda_{t[i]} + \psi_{c[i]} + \epsilon_i \\ & \forall \;\; i \in \{i : |d_{j[i]}| \leq bw\}, \end{aligned} \tag{2.3}$$

where $G_{j[i]}$ again denotes whether a given community j belongs to the German-speaking part of the country. The dependent variable $N_{j[i]}$ reflects the local norm towards private engagement prevailing in municipality j. As discussed in Section 2.3, $N_{j[i]}$ is constructed such that higher values are associated with a stronger support of private engagement; $N_{j[i]}$ reflects the share of votes rejecting more/additional government intervention. We therefore expect that $\widehat{\pi}_1 > 0$ because individuals living in the German-speaking regions are expected to show stronger support for private engagement, but also that $\widehat{\pi}_1 < 1$ because there are no sharp discontinuities in attitudes at the language border.

An additional issue in this context is that the estimated size of $\widehat{\pi}_1$ likely depends on the choice of vote(s) used to measure local attitudes towards the role of the state. We will take up this issue in the empirical analysis below, showing 2SLS estimates using different sets of votes to measure citizens' attitudes towards the role of the state. Another issue to be considered at this point is that the dependent variable in Equation (2.3) is only available at the municipality level, inducing clustering at this level (i.e. all firms within the same municipality are exposed to the same local norm). Given that we report standard errors that are clustered at the level of the community (because a firm's distance from the language border is also only available at the community-level, as mentioned above), this issue is simultaneously addressed.

Finally, we estimate the following structural equation, which relates firms' training decisions with local attitudes towards the state at the language border, using two-stage least squares (2SLS):

$$\begin{aligned} T_i = {} & \beta_0 + \beta_1 N_{j[i]} + \beta_2 X_i + \beta_3 Z_{r[i]} + \lambda_{t[i]} + \psi_{c[i]} + \epsilon_i \\ & \forall \;\; i \in \{i : |d_{j[i]}| \leq bw\}. \end{aligned} \tag{2.4}$$

Specifically, in estimating the equation above, we instrument $N_{j[i]}$ with the language-cultural background of a community using the first-stage regression from Equation (2.3) above (i.e. we use $G_{j[i]}$, indicating whether a given community belongs to the German-speaking part of Switzerland, to instrument for communal attitudes towards the role of the state, $N_{j[i]}$). The resulting 2SLS estimate of β_1 corresponds to the language-border contrast in the training intensity among employers within the discontinuity sample while also taking into account the lack of sharp discontinuity in attitudes towards the role of the state (i.e. the 2SLS estimate of β_1 represents a fuzzy regression discontinuity estimate, which in turn can be understood as a rescaled version of the reduced form estimate α_1 from Equation (2.2) above; see for example Angrist and Pischke (2008)). Given that we expect both α_1 and π_1 to be positive, it follows that we expect that $\beta_1 > 0$ as well.

2.5 Results

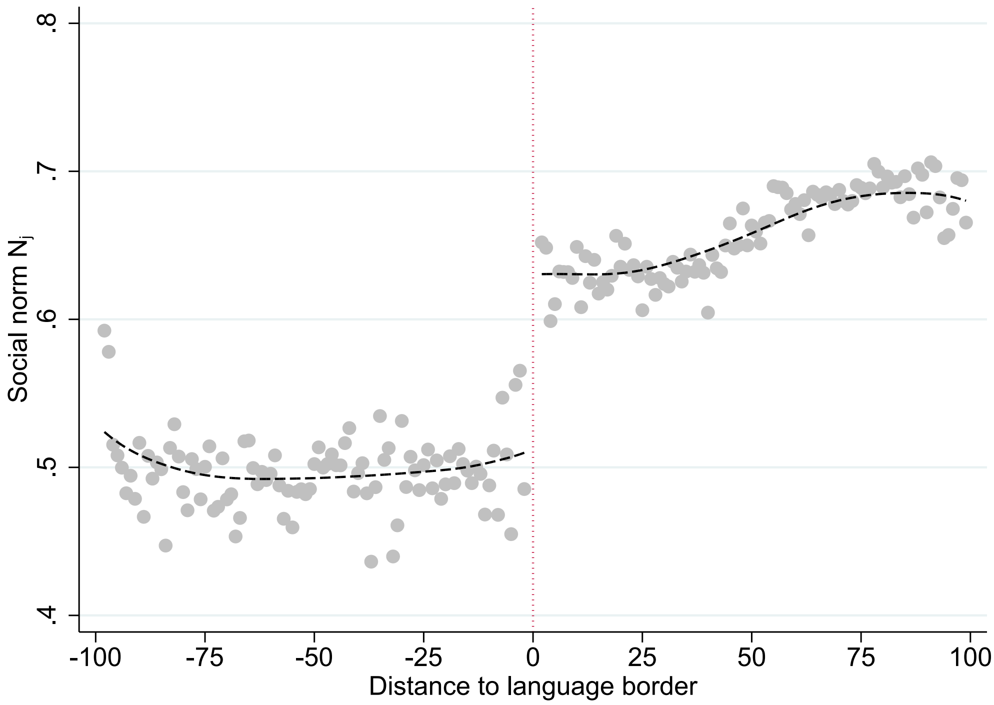
2.5.1 Graphical evidence

Moving on to our results, we start with a graphical description of the discontinuities in the local norm favoring private engagement and in the share of training firms at the language border.

First, Figure 2.3 shows that there is, as expected, a very clear and sizable discontinuity in the local norm favoring private engagement exactly at the language border (Appendix Figure B.3 shows the corresponding discontinuities for each of the eight underlying plebiscites). On average, individuals living in the German-speaking municipalities are much more supportive of private engagement than those residing in the Romance-speaking part of the country. Also note that the mean of $N_{j[i]}$ is relatively constant on both sides of the language border, consistent with the idea that the two language regions are characterized by different, relatively persistent norms regarding the question on how responsibilities should be shared between private and public actors (see also Appendix Figure B.4).

Analogously, Figure 2.4 shows the discontinuity in the share of training firms at the language border. A first thing to note is that the share of training firms over the full sample amounts to 25.7% in the Romance-speaking regions

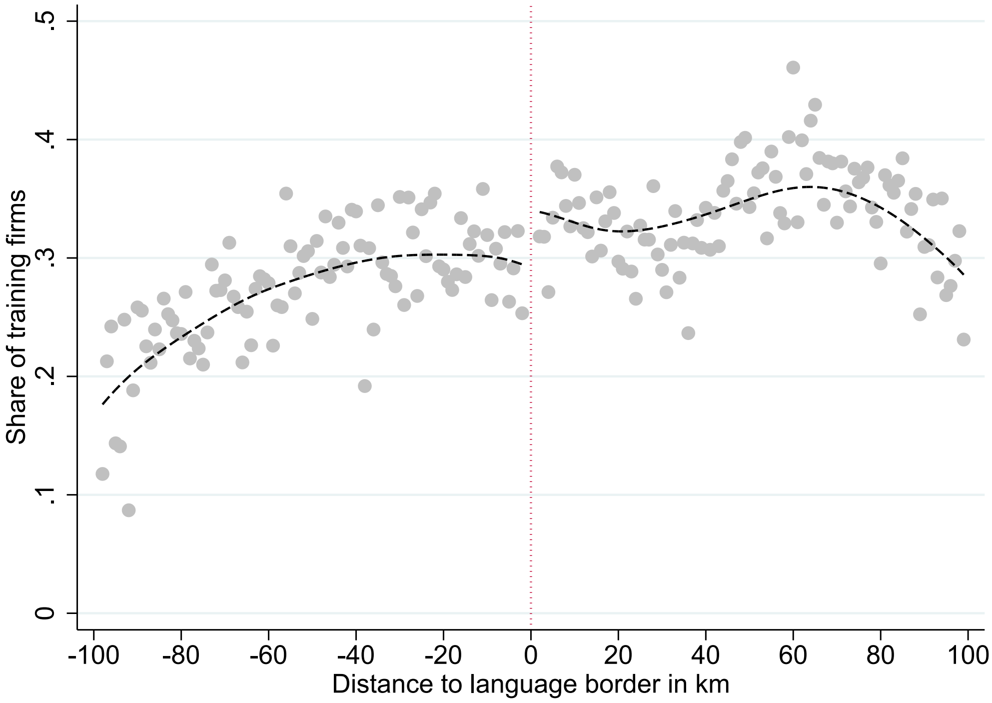
Figure 2.3: Discontinuity in the local norm favoring private engagement at the language border



Notes: The figure plots mean values of N_j for municipalities aggregated within bins of 1km width by their distance from the language border in actual travelling distance. German- (Romance-) speaking regions are associated with positive (negative) travelling distances. The dashed line shows smoothed values from a locally weighted regression.

and to 31.9% in the German-speaking regions. As is also evident from Figure 2.4, the gap in the training probability narrows somewhat when approaching the language border. Within 20 kilometers of the language border, 30.2% of firms located on the Romance-speaking side of the language border train apprentices, while 34.2% of the firms do so on the German-speaking side. Second, the shapes of the distance trends displayed on either side of the language border are ambiguous: while the overall concave pattern on the Romance side becomes more linear close to the language border, the higher training commitment among firms in German-speaking regions dips around 30 to 40 kilometers and again around 90 and 100 kilometers distance from the language border. It turns out that these patterns are mainly driven by

Figure 2.4: Discontinuity in the share of training firms at the language border



Notes: The figure shows the share of training firms aggregated within bins of 1km width by their distance from the language border in actual travelling distance. German (Romance) speaking regions are associated with positive (negative) travelling distances. The dashed line shows smoothed values from a locally weighted regression.

metropolitan areas, which are situated within these specific distance ranges and which attract large numbers of firms operating in industries where the share of training firms tends to be low.[21]

[21]Specifically, Bern is located 31 km away from the language border on the German-speaking side of the language border, Basel 37 km, Lucerne 86, and Zurich 99 km. In the Romance language regions, Lausanne (68 km), Lugano (87 km), and Geneva (128 km) contribute to the decline in training firms at greater distances from the language border. See also Appendix Figure A.1.

2.5.2 Main estimates

First-stage estimates

Moving on to our main estimates, Table 2.3 first shows the estimated discontinuity in $N_{j[i]}$ at the language border. More specifically, the table shows estimates of parameter π_1 from Equation (2.3), using the RD sample only and an expanding set of control variables. The specification shown in the first column does not include any controls and yields an estimate of $\widehat{\pi}_1 = 0.04$, with a robust standard error of 0.011. There is thus a large relative difference in the norm favoring private engagement between the two language regions of about 31% (= 100% $\cdot$ (0.147/0.474)). As expected, and consistent with the results from Eugster *et al.* (2017) among others, individuals in the German-speaking part of Switzerland are much more in favor of private engagement than those from the Romance-speaking regions.

Table 2.3: First-stage estimates (OLS estimates)

	Local norm favoring private engagement, $N_{j[i]}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$G_{j[i]}$	0.147***	0.129***	0.129***	0.129***	0.123***	0.126***
	(0.015)	(0.012)	(0.012)	(0.012)	(0.011)	(0.011)
Bandwidth	20	20	20	20	20	20
Cantonal dummies	No	Yes	Yes	Yes	Yes	Yes
Census–year dummies	No	No	Yes	Yes	Yes	Yes
Firm characteristics	No	No	No	Yes	Yes	Yes
Location characteristics	No	No	No	No	Yes	Yes
Demand controls	No	No	No	No	No	Yes
R-squared	0.643	0.821	0.821	0.823	0.873	0.880
Observations	69,619	69,619	69,619	69,619	69,619	69,619

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipality.

The remaining columns of Table 2.3 show that the estimate of π_1 remains broadly stable when we take various sets of control variables into account (we discuss the different sets of controls in some detail below when discussing the fuzzy RD estimates). Overall, the estimates from Table 2.3 are thus consistent with Figure 2.3 (which, however, uses municipality-level data; therefore the results do not exactly match with each other) and confirm that there is a substantial and statistically significant discontinuity in the local norm favoring private engagement exactly at the Swiss language border.

Reduced-form estimates

We next check whether there is a discontinuity in the share of training firms at the language border as well. With that aim, Table 2.4 presents reduced-form estimates comparing the share of training firms on either side of the language border (i.e. the table shows estimates of parameter α_1 from Equation (2.1) and (2.2), respectively). As above, we start with a model without any controls, and then we expand the set of control variables included step by step.

Column (1) of Table 2.4 shows that the share of training firms is about four percentage points higher on the German-speaking side of the language border. This implies a relative difference in the probability of training apprentices of about 13.2% ($= 100\% \cdot (0.04/0.302)$) at the language border. Again, the remaining columns of Table 2.4 show that the estimate of α_1 also remains robust to the inclusion of various sets of control variables. Thus, the estimate of parameter α_1 remains pretty much the same whatever the set of controls, as expected when the observable characteristics on either side of the discontinuity are balanced (e.g. Lee and Lemieux, 2010).

Fuzzy RD estimates

However, Figure 2.3 and the estimates from Table 2.3 also show that there is no sharp discontinuity in the local norm favoring private engagement; the estimated discontinuity in $N_{j[i]}$ is much smaller than 100 percent. To take this feature into account, we next estimate a fuzzy RD model, in which we

Table 2.4: Reduced form RD-estimates (OLS estimates)

	Training firm, T_i					
	(1)	(2)	(3)	(4)	(5)	(6)
$G_{j[i]}$	0.040***	0.046***	0.046***	0.044***	0.043***	0.045***
	(0.011)	(0.012)	(0.012)	(0.011)	(0.010)	(0.010)
Bandwidth	20	20	20	20	20	20
Cantonal dummies	No	Yes	Yes	Yes	Yes	Yes
Census–year dummies	No	No	Yes	Yes	Yes	Yes
Firm characteristics	No	No	No	Yes	Yes	Yes
Location characteristics	No	No	No	No	Yes	Yes
Demand controls	No	No	No	No	No	Yes
R-squared	0.002	0.004	0.005	0.122	0.124	0.125
Observations	69,619	69,619	69,619	69,619	69,619	69,619

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipality.

instrument $N_{j[i]}$ with $G_{j[i]}$, a dummy variable indicating whether municipality j belongs to the German-speaking part of Switzerland.

Column (1) of Table 2.5 shows the estimate from a first specification without any controls. The resulting point estimate of $\widehat{\beta}_1 = 0.27$ yields a positive relation between our norm measurement $N_{j[i]}$ and firms' training propensity. Moreover, the estimate is statistically significant at the 1% level.

The remaining columns of Table 2.5 show that this estimate is robust to the inclusion of several sets of controls. Column (2) and column (3) add dummies for canton and census year, respectively. Again, including the cantonal dummies changes the interesting parameter estimate somewhat, suggesting that institutional differences across cantons are relevant, but including census-year dummies has no discernible impact on the size of $\widehat{\beta}_1$.

Next, column (4) further adds firm-level controls, such as firm size and industrial affiliation. The inclusion of the firm covariates hardly changes the point estimate of β_1, which remains of similar size and statistically significant;

Table 2.5: Fuzzy RD-estimates (2SLS estimates)

	Training firm, T_i					
	(1)	(2)	(3)	(4)	(5)	(6)
$N_{j[i]}$	0.270***	0.358***	0.359***	0.344***	0.350***	0.355***
	(0.077)	(0.106)	(0.106)	(0.095)	(0.093)	(0.087)
Bandwidth	20	20	20	20	20	20
Cantonal dummies	No	Yes	Yes	Yes	Yes	Yes
Census year dummies	No	No	Yes	Yes	Yes	Yes
Firm characteristics	No	No	No	Yes	Yes	Yes
Location characteristics	No	No	No	No	Yes	Yes
Demand controls	No	No	No	No	No	Yes
R-squared	0.001	0.003	0.003	0.121	0.123	0.125
Observations	69,619	69,619	69,619	69,619	69,619	69,619
F-statistic (first-stage)	92.23	116.7	116.7	118.2	134.1	124.0

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipality.

the inclusion of firm-level controls decreases the standard error associated with $\widehat{\beta}_1$. Next, in column (5), we add several locational controls. Again, this leads to a somewhat smaller standard error without much impact on the size of the point estimate.

Finally, column (6) accounts for factors shifting the demand for apprenticeships and displays our baseline specification. As outlined in Section 2.3, we struggle to fully disentangle firms' supply of apprenticeships under discussion from youngsters' demand for apprenticeships. Including exogenous apprenticeships demand shifters in column (6), namely the minimum distance to any upper-secondary schooling type and the share of the populace between the age of 15 and 25, mitigates this concern. Compared to the previous specifications, the point estimate remains fairly stable at 0.36.

All specifications considered, the estimates shown in Table 2.5 suggest a positive association between a locally persistent norm favoring private provision of goods and the training propensity of locally present firms. This finding is not sensitive to the step-by-step inclusion of firm characteristics, locational characteristics, or controls for variation in the demand for apprenticeship training.

It is also worth comparing the estimate in column (6) with those from Kuhn *et al.* (2019), who report a baseline estimate of 0.393 (as well as somewhat larger 2SLS estimates). Since the norm measurements are calculated very similarly and the outcome variable T_i is virtually the same in both studies, this estimate can be directly compared to our baseline estimate of 0.355. Thus, the two identification approaches yield estimates that are very close to each other, and therefore also consistent with each other, even though they are based on different data sources (and thus different samples of firms) and different empirical approaches. This further corroborates our results.

2.5.3 Robustness

We next provide several alternative specifications and subsample estimations to ensure the robustness of our results.

Alternative specifications

We start with some alternative specifications, shown in Table 2.6 (for ease of comparison, the first column replicates the baseline specification from column (6) of Table 2.5).

As a first robustness check, we use a bandwidth of 50 (instead of 20) kilometers around the language border, which increases the number of observations substantially. Compared to our baseline estimate, however, this hardly changes the resulting estimate of interest ($\widehat{\beta}_1 = 0.31$). As expected, though, the estimate in column (2) is more precisely estimated than the baseline estimate. A second check includes the travelling distance from the language border, $d_{j[i]}$, and the interaction between $d_{j[i]}$ and $G_{j[i]}$, as additional control variables (cf. column 2 of Table 2.6). This again yields a comparable

Table 2.6: Robustness (2SLS estimates)

	Training firm, T_i							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline	50 km bandwidth	Including $d_{j[i]}$ and $d_{j[i]} * G_{j[i]}$	Including micro firms	Within MS-regions	$N_{j\|i}$ = votes in sample period	$N_{j\|i}$ = non-VET votes	$N_{j\|i}$ = VET votes
$N_{j[i]}$	0.355***	0.310***	0.429***	0.164***	0.348*	0.351***	0.315***	0.572***
	(0.087)	(0.065)	(0.157)	(0.062)	(0.209)	(0.085)	(0.076)	(0.158)
Bandwidth	20	50	20	20	20	20	20	20
Cantonal dummies	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
LM-region dummies	No	No	No	No	Yes	No	No	No
Census year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demand controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.125	0.103	0.125	0.119	0.126	0.125	0.125	0.124
Observations	69,619	294,021	69,619	121,429	69,619	69,619	69,619	69,619
F-stat(first-stage)	124.0	294.9	64.45	136.7	39.89	127.9	138.4	58.85

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipality.

estimate of $\widehat{\beta}_1 = 0.429$.[22]

Column (3) reports the resulting estimate of β_1 when we also include the smallest firms in our sample (those with less than three employees). This leads to a large increase in the sample size and, because these firms are much less likely to train any apprentices, to a reduction in the point estimate of β_1. However, the relative size of the estimate remains about the same as in the baseline specification. To illustrated this, one can compare the estimates in columns (1) and (4) with the share of training firms in the two samples of 0.32 and 0.20, respectively.

As a next robustness check, the specification shown in column (5) includes fixed effects at the level of local labor markets instead of cantonal-level fixed effects.[23] The resulting estimate is fairly close to the baseline estimate, suggesting that our results are not driven by differences in product demand or labor supply that the firms face at the level of local labor markets. However, the point estimate is less precisely estimated than those in the other columns. This comes as no surprise, as only nine out of 106 MS regions are bilingual and only three MS regions have a linguistic minority of at least a quarter of the total population.

Finally, the last three columns of Table 2.6 report estimates that use slightly different parameterizations of our norm measurement, $N_{j[i]}$. The construction of $N_{j[i]}$ used in column (6) uses only the results from votes 458 to 543, such that the time period implicitly covered by $N_{j[i]}$ is closer to that in the Business Census. In column (7), we only use the results from the votes not directly related to VET policies (i.e. all votes except 340 and 503; cf.

[22] In Appendix A, we provide a more thorough analysis related to the choice of bandwidth and the inclusion of a distance term.

[23] Local labor markets correspond to the MS regions defined by the Federal Statistical Office (where MS stands for spatial mobility in French, *mobilité spatiale*). While cantons are arguably Switzerland's most important institutional entities, labor markets do not often coincide with these institutional borders. MS regions are therefore defined by the FSO as regions with common commuting patterns towards their centers. We thus argue that the coefficient from column (5) yields the correlation between the local norm regarding private engagement in the provision of goods and the training behavior of firms that operate in the same market in terms of products they sell and labor supply they require.

Appendix Table B.2), while column (8) is based on using these two votes only. Column (7) yields an estimate of $\widehat{\beta}_1 = 0.315$, which is again very close to our baseline estimate. In contrast, the estimate from column (8) turns out to be somewhat larger ($\widehat{\beta}_1 = 0.572$) than in the baseline specification, but the difference between the two estimates is not statistically significant.

Subsample estimates

We next turn to estimates for various subsamples in Table 2.7 (again, the first column simply replicates our baseline estimate from column (6) of Table 2.5). As above, the aim of these additional estimates is to probe the robustness of our main finding.

First, columns (2) and (3) show that the point estimate of β_1 is similar to the baseline estimate when we restrict the sample solely to either 1998 or 2008. The resulting estimates are both close to the baseline estimate, suggesting that there is no problem in pooling data from several waves of the Business Census (we find broadly similar estimates for the other two time periods; results not shown).

The next two columns show that the relation between a norm favoring private engagement and firms' training participation is persistent among both small and large firms (where small firms are defined as firms with three to 10 employees). Note that in relative terms the estimate in column (4) is actually larger than the estimate in column (5) due to different training participation rates between the two subsamples (0.26 among small firms versus 0.50 among middle and larger firms).

In column (6) of Table 2.7, we restrict the sample to for-profit firms by excluding not-for-profit and public employers. This accounts for the fact that the share of for-profit and public firms is not entirely balanced across the two language regions (see Table B.1). The resulting point estimate of $\widehat{\beta}_1 = 0.40$ is close to our baseline estimate and remains highly significant: it shows that our results are not driven by state authorities, which arguably face different budget restrictions than privately-run firms.

Next, the specification shown in column (7) focuses on a narrow subsam-

Table 2.7: Subsample results (2SLS estimates)

	Training firm, T_i							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline	Wave 1998	Wave 2008	Small firms	Middle and large firms	For-profit firms	Bilingual cantons	French/German subsample
$N_{j[i]}$	0.355***	0.371***	0.452***	0.260***	0.349**	0.400***	0.317***	0.364***
	(0.087)	(0.102)	(0.102)	(0.094)	(0.144)	(0.096)	(0.114)	(0.112)
Bandwidth	20	20	20	20	20	20	20	20
Cantonal dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Census year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demand controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.125	0.125	0.130	0.141	0.158	0.113	0.129	0.104
Observations	69,619	17,370	17,654	51,042	18,577	59,023	34,160	61,087
F-stat(first-stage)	124.0	111.7	131.9	120.2	138.2	129.6	438.3	351.6

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipality.

ple of firms located in the three bilingual Cantons of Bern, Fribourg, and Valais. As discussed in Section 2.2.2, the language border partly coincides with cantonal borders. Due to the extensive political autonomy Swiss cantons exhibit, this may threaten our identification strategy. Restricting the sample to bilingual cantons allows us to tackle this concern.[24] The resulting point estimate of $\widehat{\beta}_1 = 0.317$ is very close to our baseline estimate and relatively precisely estimated. This makes our model appear to be robust to institutionally shaped demand for apprenticeships (for additional discussion on demand-side effects see Sections 2.6.2 and 2.6.2).

Finally, as outlined in Section 2.2.2, Switzerland consists of four language regions. The two largest of these regions are the German-speaking and the French-speaking speaking regions, separated by the so-called *Röstigraben* (named after the Swiss-German potato dish *Rösti* and the German word for rift, *Graben*). In a final subsample specification, we focus only on this specific part of the language border, which again reduces the sample size somewhat. Column (8) in Table 2.7 suggests that our results are not driven by other minor language borders within Switzerland but remain persistent when evaluated solely at the *Röstigraben*.

2.6 Additional evidence on mechanisms

In the final section of the empirical analysis, we provide evidence supporting our argument that local norms partially determine training behavior among employers, and we also try to rule out some alternative explanations.

2.6.1 Norm internalization

A norm favoring private provision of common goods should, not only be visible in voting results but also in other aspects of peoples' lives and through their actions. Obviously, it is hard to differentiate whether any action, preference, or belief in line with a norm contributed to the norm's evolution in

[24]Note: In addition to the educational system, cantons determine other labor market relevant policies, such as personal and corporate taxation and unemployment schemes.

the first place, or whether the norm was the origin of the action, preference, or belief. However, we argue that actions and expressed preferences and beliefs conforming to a norm demanding private provision of goods can serve as evidence for individuals' internalization of this norm. In order to evaluate this claim, we analyze four additional data sources.[25] For each outcome considered, Table 2.8 shows the contrast between the two language regions, both without and with additional controls.

First, panel (a) of Table 2.8 shows that the share of individuals that prefer state over private ownership of businesses is lower among German-speaking regions than among Romance-speaking regions (item 1). Second, panel (a) reveals that German-speaking individuals trust other people more than Romance-speaking individuals (item 2). Both differences remain widely unchanged when evaluated conditional on various individual and municipality covariates in the last column. We argue that these findings are in line with a norm valuing private provision of goods and possibly reveal two mechanisms supporting it: Individuals who are critical towards state-owned business are more likely to assume responsibility in providing goods. And larger trust towards others might mitigate concerns that peers are purely financially incentivized and thus stop contributing to a good (Fischbacher *et al.*, 2001; Kollock, 1994). In the context of VET, this lowers negative poaching externalities for firms offering apprenticeship positions.

Next, panel (b) of Table 2.8 compares three measures of individuals' private engagement across the two language regions. These three items ask whether people (3) actively participate in associations (sports, social, or cultural), (4) worked voluntarily during the last four weeks, and (5)

[25]First, the World Values Survey allows comparison of average language region differences because the only spatial information the survey contains for Swiss respondents is the language region they live in. Second, the Swiss Volunteering Survey 2006 conducted by the Swiss Society for the Common Good (SSCG) asks people about their voluntary commitment; results are available on municipality level. Third, the Swiss Household Panel is a longitudinal data set composed of two cohorts of randomly chosen Swiss households, surveyed annually. As a spatial indicator, it includes 106 Swiss labor market regions. Fourth, we use data from an exit poll related to vote 503 from the year 2003 (cf. Appendix Table B.2).

Table 2.8: Evidence on norm internalization

	Question	N	Mean answer: Romance	German	Difference	Conditional difference[a]
(a) Trust						
(1)	"State vs. private ownership of business": WVS 1994/98 and 2005/09					
		2,249	4.263	3.884	-0.379***	-0.382***
			(0.077)	(0.063)	(0.099)	(0.099)
(2)	"Trust other people": SHP 1999-2012[b]					
		3,892	5.440	6.224	0.784***	0.574***
			(0.063)	(0.045)	(0.076)	(0.084)
(b) Voluntary commitment						
(3)	"Membership in associations": Swiss Volunteering Survey 2006/10					
		4,648	0.664	0.775	0.111***	0.120***
			(0.012)	(0.007)	(0.014)	(0.027)
(4)	"Voluntary work during last 4 weeks": Swiss Volunteering Survey 2006					
		2,753	0.412	0.512	0.100***	0.093*
			(0.016)	(0.012)	(0.020)	(0.039)
(5)	"Donated during last year": Swiss Volunteering Survey 2006/10					
		4,648	0.721	0.832	0.112***	0.066**
			(0.012)	(0.007)	(0.013)	(0.025)
(c) Voting motivation ("apprentice initiative")						
(6)	"Reason for 'No': responsibility of firms/no state task": VOX 2003					
		326	0.289	0.464	0.175**	0.183***
			(0.052)	(0.032)	(0.064)	(0.065)

Notes: Sample restricted to observations within 50 kilometers of the language border; due to a lack of municipality identifiers, this restriction was not possible in (1) and (6). Panel A, question (1) refers to a ten point scale ranging from 1: "Private ownership of business should be" to 10: "Government ownership of business should be". Panel A, question (2) refers to a 0 ("One can't be careful enough") to 10 ("One can trust most people") scale to the question "Would you say, one can trust most people?". Answers displayed in panel B are binomial Yes/No. Panel C refers to a 1 (No-votes in vote 503 (see B.2) indicated as reason for their individual vote either "economy should react by itself" or "firms not responsible for (apprenticeship-)market" or "can't be forced by the state" or "other statement concerning the responsibility of the economy" or "self-responsible" and a 0 otherwise. [a]OLS regression of answers on a German language dummy, demographic and municipality (not included in question (1) of panel A due to a lack of municipality identifier) controls. [b]Though the Swiss Household Panel has a panel structure, we only include the latest observation per person.

donated during the last year. All three indicators display higher values for individuals in German-speaking regions and are again robust to the inclusion of covariates in the last column. Overall, financially unrewarded private commitment seems to be more widespread among German-speaking individuals. This appears to be in line with a norm demanding the private provision of goods.[26]

Finally, panel (c) uses data from an exit poll related to vote 503 (cf. Appendix Table B.2), which explicitly called for more state responsibility within the apprenticeship system. We here focus on the motivation of no-voters.[27] Item (6) shows that 29% of all no-voters in Romance-speaking regions stated that the provision of apprenticeships is not a task the state is responsible for, while the equivalent share amounted to 46% among no-voters in German-speaking regions.[28] This directly supports the claim that German speakers are more critical of state engagement in the provision of apprenticeship positions than their Romance-speaking counterparts.

Altogether, we conclude that German-speaking individuals mistrust government-owned business more, whereas they display higher trust towards others and show higher voluntary and financially unrewarded engagement. Moreover, they are more critical of state engagement in the provision of apprenticeship positions. We believe that these findings correspond to a norm demanding private engagement in the provision of goods in two reciprocal ways. First, this norm might be rooted simultaneously in individuals' mistrust of government-owned business and their intensive involvement within their local communities. Secondly, these attitudes and actions sustain this norm once it occurs.

[26] See also Appendix Figure B.5 which again highlights differences in the share of people working voluntarily at the language border

[27] Table 2.2 exploits the same data source to validate our norm measurement.

[28] In contrast, Table 2.2 documents no differences across language regions in the motivation of yes-voters to approve the vote due to a perceived lack of apprenticeship positions.

2.6.2 Selection into apprenticeship training, expected labor market outcomes, and apprentice wages

We next focus on alternative explanatory mechanisms and try to rule out several alternative mechanisms as far as possible.

Selection into apprenticeship training

A first mechanism that could threaten our identification strategy is differential selection of pupils from the two language regions into apprenticeship training and other educational tracks, namely high school.[29] More specifically, given the number of potential apprentice applicants, e.g. pupils with sufficient cognitive abilities, pupils preferring high school over apprenticeship lower the number of applicants for apprenticeships and may consequently decrease the number of apprenticeship positions filled by firms. If these preferences contrast at the language border, we might wrongly interpret lower numbers of apprentices as lower apprenticeship provision by firms.

In order to identify whether selection into apprenticeship and high school differs between the two language regions, we use PISA data from the bilingual canton of Bern.[30]

Column (1) of Table 2.9 shows the estimates of running a simple regression of pupil i's PISA score on a German-speaker dummy G_i, a dummy A_i taking the value 1 (zero) if pupil i plans to opt for an apprenticeship (for high school) after the summer break, and their interaction term $G_i \times A_i$.[31] We are mainly interested in the coefficient associated with the interaction term, as

[29] In a survey of 5,934 adults, Cattaneo and Wolter (2016) find that apprenticeship training has lower prestige among Romance speakers than German speakers. This raises concerns about diverging selection mechanisms across the language border.

[30] The Canton of Bern is the only bilingual canton covered by the PISA survey and with observations from both language regions. We thus focus solely on this canton to be able to hold the institutional setting constant. We pool data from the 2000 and 2009 waves to increase the sample size.

[31] We only consider pupils planning to opt either for apprenticeship or for high school. This approach is inspired by Bolli and Rageth (2016), who compare selection into apprenticeship training between native and immigrant children.

Table 2.9: Selection into apprenticeship training, labor market outcomes, and apprentice wages (OLS estimates)

	PISA score	ln(wage)		Unemployment		ln(apprentice wage)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
G_i	37.1***	0.057***	-0.005	-0.024***	-0.010	0.047***	0.035
	(3.90)	(0.004)	(0.008)	(0.002)	(0.010)	(0.010)	(0.039)
A_i	-70.4***	-0.296***	-0.259***	-0.012***	-0.007***		
	(4.44)	(0.004)	(0.004)	(0.002)	(0.002)		
$G_i \times A_i$	-8.10	-0.006	-0.005	0.008***	0.005*		
	(5.30)	(0.005)	(0.005)	(0.003)	(0.002)		
Data source	PISA	SLFS		SLFS		SLFS	
Subsample	Canton of Bern	Working population, ages 20-60		Total population, ages 20-60		Apprentices	
Cantonal dummies	No	No	Yes	No	Yes	No	Yes
Survey-year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demographic controls	Yes	No	Yes	No	Yes	No	Yes
Sectoral dummies	No	No	Yes	No	Yes	No	No
Occupational dummies	No	No	No	No	No	No	Yes
Observations	2,118	137,679	137,679	145,474	145,474	12,195	12,195
R^2	0.422	0.116	0.371	0.061	0.139	0.005	0.179

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses. Dependent variables are (1) PISA scores (average of math, reading, and science), log monthly earnings per FTE (2/3), an unemployment dummy (4/5), and log apprentice wage (6/7). A_i refers to a dummy taking the value 1 for pupils planning to start an apprenticeship (in column 1) and a dummy takeing the value 1 one for people who entered an apprenticeship after compulsory schooling (in columns 2-5).

a positive interaction term would indicate a stronger selection of relatively high-ability pupils into VET among German-speaking pupils than among Romance-speaking pupils. However, the interaction term is not significantly different from zero. We thus conclude that, given pupils' cognitive ability, selection into one or other track after compulsory school is similar on either side of the language border.

Expected labor market outcomes

Another channel potentially shaping the demand for apprenticeships are expected labor market outcomes typically associated with this educational track. We therefore use additional data from the Swiss Labor Force Surveys 2010-2014 to link adults' first completed postcompulsory education with their contemporary earnings and working status and focus on differences between the language regions.

In columns (2) to (5) of Table 2.9 we thus check whether labor market outcomes for the two groups differ between the two language regions. Similar to column (1), we run a simple regression of individual i's labor market outcome on a German-speaker dummy G_i, a dummy A_i with 1 if individual i opted for an apprenticeship and 0 if individual i opted for high school or another general track after compulsory schooling, and their interaction term $G_i \times A_i$. Also, we are again mainly interested in the coefficient associated with the interaction term $G_i \times A_i$.[32] This term discloses whether relative labor market differences between people initially opting for apprenticeship or high school vary among language regions. While this is not the case for wages (columns 2 and 3), row 3 of columns (4)

[32] The estimates of the German-speaker dummy G_i in row 1 display general differences between German- and Romance-speaking regions which are not statistically significant when including demographic, sector, occupation, and cantonal controls in columns (3) and (5). According to the estimates of the apprenticeship dummy A_i in row 2 of column (2), individuals who initially opted for an apprenticeship track earn less than their counterparts who initially opted into high schools, but they are 1.2 percentage points less likely to be unemployed (row 2, column 4). These differences persist when including controls (row 2, columns 3 and 5).

and (5) suggest that the relative apprenticeship advantage over high-school graduates in unemployment rates is 0.5 and 0.8 percentage points larger in Romance-speaking regions than in German-speaking regions, respectively. Thus, expected labor market outcomes should at least not shape demand for apprenticeships more negatively among Romance-speaking pupils than among German-speaking pupils.

Apprentice wages

We further analyze differences in apprentice wages between the two language regions in the Swiss Labor Force Surveys 2010-2014. Because apprentice wages are set relatively competitively in Switzerland, we assume they partly reflect demand for and supply of apprenticeships. A priori, it is ambiguous in which language region apprentice wages are higher. A higher supply of apprenticeships by firms for a given number of apprenticeship applicants across German-speaking regions would ceteris paribus be expressed by higher apprentice wages in German-speaking regions than in Romance-speaking regions. However, high numbers of applicants for firms' given supply of apprenticeships in German-speaking regions would ceteris paribus lead to lower apprentice wages among German-speaking regions.[33]

Column (6) of Table 2.9 documents 4.7 percent higher apprentice wages in German-speaking regions than in Romance-speaking regions. This difference decreases to 3.5 percent once we include demographic controls and the VET occupation and, moreover, appears to be statistically insignificant (column 7). Cautiously, we surmise that somewhat higher apprentice wages among firms in German-speaking regions reflect a higher initial supply of apprenticeships by firms across these regions.

[33]Note that in light of lower shares of apprentices across Romance-speaking regions, we rule out the possibility of a higher supply of or demand for apprenticeships in Romance-speaking regions than in German-speaking regions.

2.6.3 Number of applications per apprenticeship position

Perhaps the most compelling piece of empirical evidence in support of our claim that the discontinuity in the training probability among employers is not only driven by the pupils demand for apprenticeships comes from an additional analysis using data from the fourth survey on the costs and benefits of apprenticeship training (Gehret *et al.*, 2019). In the survey, among many other things, employers were asked about the number of applications they receive for each open apprenticeship position.

Table 2.10: Log number of applications per apprenticeship position (OLS estimates)

	ln(number of applications)			
	(1)	(2)	(3)	(4)
G_j	-0.180***	-0.366***	-0.482***	-0.360***
	(0.033)	(0.075)	(0.085)	(0.137)
Bandwidth	*full sample*	*full sample*	20	20
Cantonal dummies	No	Yes	No	Yes
Firm characteristics	No	Yes	No	Yes
Location characteristics	No	Yes	No	Yes
Demand controls	No	Yes	No	Yes
R^2	0.005	0.393	0.051	0.576
Observations	5531	5531	596	596

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses.
Source: Gehret *et al.* (2019), own calculations.

Table 2.10 shows that employers in the Romance-speaking regions receive more applications on average than those in the German-speaking regions (see also Appendix Figure B.6). The difference in the number of applications is large and statistically significant, both in the full and the RD sample. In the full sample, the estimated discontinuity in the number of applications received is about 18% to about 36.6% lower among employers located on the German-speaking side of the language border, depending on whether controls are taken into account or not. Similarly, in the RD sample, the estimated discontinuity at the language border is about -48.2% without to about -36% with controls. Assuming that there is no difference in search behavior across the two language regions, these estimates suggest that the demand for apprenticeship training (relative to the supply) is higher, not lower, in the Romance-speaking part of the country.

2.6.4 Alternative educational tracks

Beside apprenticeships, high schools are the most relevant track at the upper-secondary level. At the national level, roughly 20% of all pupils enter this track annually. Firms recruiting apprentices after compulsory schooling thus compete to some extent with high schools for high-ability pupils. High school rates vary substantially across Swiss cantons and tend to be higher among Romance-speaking cantons.[34] Therefore, one may conclude that firms in Romance-speaking regions simply do not train because they lack potential apprentices. However, this subsection examines two alternative tracks for pupils after compulsory school that could instead provide firms with apprenticeship applicants. Both these tracks are more common among Romance-speaking pupils.

[34]The share of pupils opting for high schools after compulsory schooling varies from the Cantons of Obwalden (11.0%), Glarus (12.2%), and Schaffhausen (13.0%) at the lower end to Ticino (27.3%), Geneva (29.4%), and Basel-Stadt (29.6%) at the upper end in 2016; source: https://www.bfs.admin.ch/bfs/en/home/statistics/catalogues-databases/tables.assetdetail.2421478.html.

School-based apprenticeship

The present chapter focus on firms' provision apprenticeship positions in dual apprenticeship programs.[35] However, pupils opting for an apprenticeship after compulsory schooling can also enter school-based apprenticeship programs, which are publicly financed and taught at special schooling facilities. While these programs equip youngsters with similar vocational skills as provided during a dual apprenticeship program in those occupations where both exist, they do not rely on the participation of firms. This makes school-based and dual apprenticeships close substitutes from the viewpoint of pupils.

Figure 2.5 displays apprentices attaining a school-based apprenticeship program as a share of the total apprenticeship cohort. Within 20 km of the language border, this share amounts to 14.0% in Romance-speaking regions but only 6.9% in German-speaking regions. Although we cannot rule out that Romance-speaking youngsters, and their parents, partly prefer school-based apprenticeship programs over dual apprenticeships, we argue that high numbers of school-based apprentices represent potential applicants for firms aspiring to train apprentices in Romance-speaking regions.

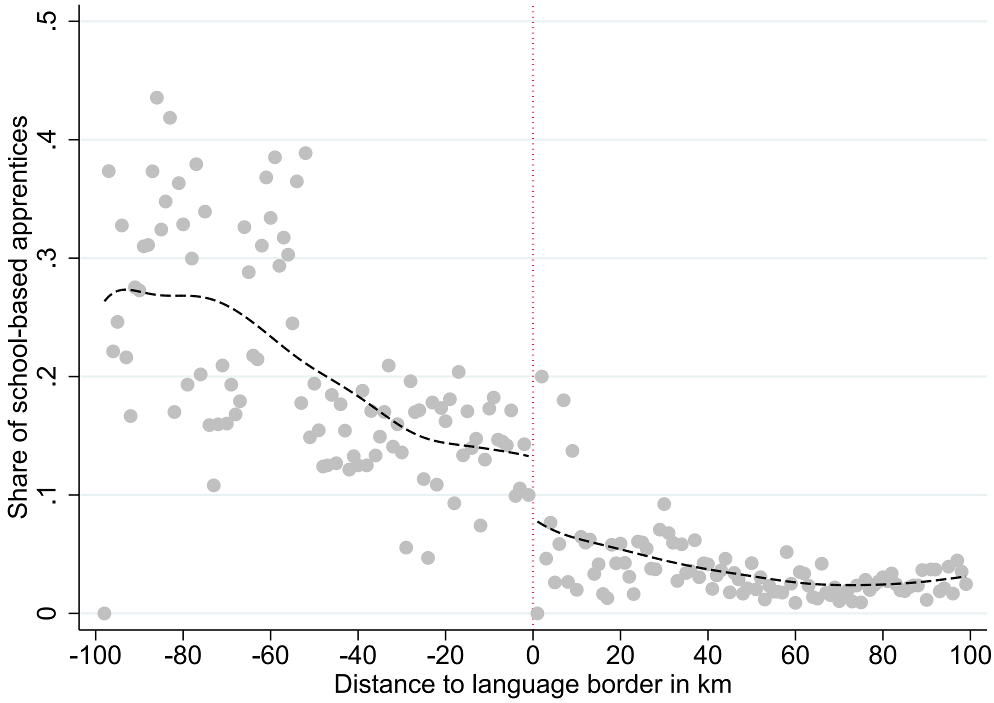
Entering the labor market without upper-secondary education/training

Another option is to enter the labor market directly after mandatory schooling, without any additional education or training at the upper-secondary level.

Panel (a) of Table 2.11 yields a higher participation in the regular labor market for people between age 16 and 18, and thus at the regular age of apprentices, in Romance-speaking regions. The statistically significant difference of 0.7 percentage points is evaluated in a 20 kilometer bandwidth of the language border and amounts to roughly 10% within this age cohort. Moreover, this difference is persistent among a Swiss-only subsample (panel b). This is somewhat surprising, considering that this is reversed for older

[35]The term *dual* refers to the fact that apprentices in these programs spend three to four days per week in the training firm and the other one to two days per week in a vocational school.

Figure 2.5: Discontinuity in the share of school-based apprenticeships at the language border



Notes: The figure shows the share of school-based apprentices, aggregated by the distance of their place of residence from the language border within bins of 1km width. The dashed line shows smoothed values from a locally weighted regression.
Source: Federal Statistical Office, Vocational Education and Training (VET) – Apprenticeships 2016.

cohorts, as panels (c) and (d) display. As in the case of school-based apprentices, we argue that these youngsters entering the regular labor market directly after compulsory school represent potential applicants for firms aspiring to train apprentices in Romance-speaking regions.

2.6.5 Religion

A final alternative explanation is religion. Switzerland is not only linguistically diverse but also heterogeneous in terms of individuals' religious affiliation. Although language regions are not religiously homogeneous, they

Table 2.11: Employment rates for different age cohorts (OLS estimates)

	Full	RD sample			
Age group	sample	All	Romance	German	Difference
(a) Ages 16-18					
	0.070	0.070	0.074	0.066	-0.007**
	(0.001)	(0.002)	(0.003)	(0.002)	(0.003)
	[235,437]	[22,942]	[10,692]	[12,250]	
(b) Ages 16-18, Swiss only					
	0.058	0.061	0.065	0.057	-0.008**
	(0.001)	(0.002)	(0.003)	(0.002)	(0.003)
	[196,610]	[19,885]	[8,992]	[10,893]	
(c) Ages 19-25					
	0.541	0.515	0.455	0.578	0.123***
	(0.001)	(0.002)	(0.003)	(0.003)	(0.004)
	[542,391]	[52,585]	[26,955]	[25,630]	
(d) Ages 30-55					
	0.599	0.590	0.590	0.602	0.012***
	(0.000)	(0.001)	(0.001)	(0.001)	(0.002)
	[2,537,388]	[246,513]	[114,674]	[131,839]	

Notes: The table shows employment rates in the full and the RD sample for different age groups. The square brackets contain the number of observations in the respective cohort. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively.
Source: Population Census 2000.

differ on average: 55% of the people living in Romance language regions are Catholic, and only 21.3% Protestants. In contrast, these shares are almost reversed in the German-speaking region, with 37.8% Catholics and 42.1% Protestants (cf. Appendix Figure B.7). If religious affiliation correlates with, or even causes, private engagement, an imbalance in religious affiliation at the language border could threaten our identification strategy. Indeed, Basten and Betz (2013), who apply a spatial RDD at the border between the Cantons of Fribourg (mainly Catholic) and Vaud (mainly Protestant), find stronger preferences for leisure, redistribution, and government intervention among Catholics. However, differing institutional frameworks across cantons, including educational institutions, mean that a cantonal border is not very well suited for evaluating the association between prevalent norms and the provision of apprenticeship positions (Eugster *et al.* (2017) proposed a similar argument).

To test for the influence of religion, we thus exploit another natural laboratory existing in Switzerland: the religiously divided but entirely German-speaking Canton of Aargau. The Canton of Aargau lies in the northeast of Switzerland, covers 3.4% of its territory, and accounts for 7.8% of its total population. Short distances and common institutions make the Canton of Aargau a well-integrated labor market, which remains, however, religiously separated for historical reasons.[36]

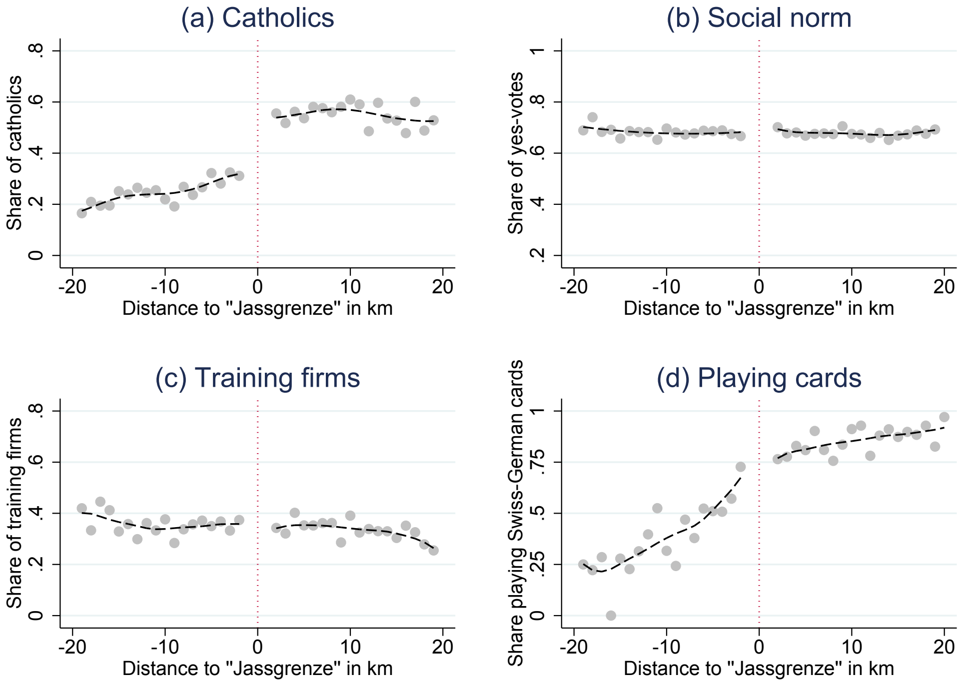
In the contemporary Canton of Aargau, this religious border is known as the *Jassgrenze*, and panel (a) of Figure 2.6 shows that it is still visible.[37] Across the territories formerly ruled by the Bernese, only 24.5% declare themselves to be Catholics, but 54.2% do so in the rest of the canton. Reassuringly for our analysis, these different shares of Catholics neither translate into diverging levels of our norm measurement $N_{j[i]}$ nor into differences in the share of training firms along the *Jassgrenze*, as shown in panels (b) and (c) of Figure 2.6, respectively.[38]

[36]Before its founding in 1803, the canton was split between the city-state of Bern and the Old Swiss Confederacy. During the Reformation in the 16th century, the Bernese rulers imposed Protestantism on their territories; the rest territories remained Catholic (Seiler and Steigmeier, 1991).

[37]Named after the Swiss card-game *Jass* and the German word for border (*Grenze*) due to the use of different sets of playing cards on either side of this border (panel d of Figure 2.6). The border is also known as the Brünig-Napf-Reuss line, named after the Alpine pass Brünig, the mountain Napf, and the river Reuss it crosses and follows, respectively.

[38]Table B.3 displays the corresponding estimates.

Figure 2.6: Religious affiliation, local norm, training firms, and playing cards in the Canton of Aargau



Notes: The figures show (a) the share of Catholics, (b) mean values of N_j for municipalities aggregated, (c) the share of training firms, and (d) the share of players using the German cards, aggregated by the distance from the religious border (*Jassgrenze*) within bins of 1km width. The dashed lines show smoothed values from locally weighted regressions.
Sources: (a) Population Census 2000; (b) Federal Statistical Office 2017; (c) Business Census 1998, 2001, 2005, and 2008; (d) Swisslos Interkantonale Landeslotterie, Swiss Jass Championships held online in 2017 and 2018.

2.7 Conclusions

In this chapter, we ask whether a local norm favoring more private engagement enhances firms' voluntary provision of apprenticeship positions. To evaluate this hypothesis empirically, we apply a spatial RDD at the Swiss language border, where voting results reveal remarkable differences in a norm preferring private provision of goods.

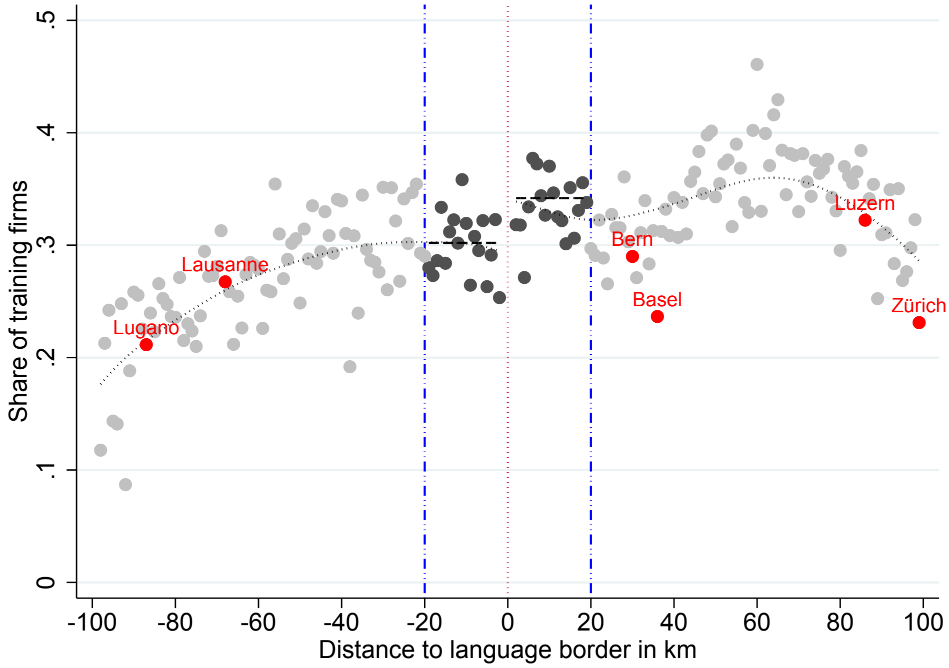
Our empirical analysis establishes two main results. First, individuals living in the German-speaking regions favor private provision of goods, while their Romance-speaking counterparts prefer the state provision of these goods. Second, our preferred estimate associates a one-standard-deviation difference in our norm measurement with a 3.6 percentage point difference in the share of training firms. Considering firms' average training propensity of 30% across all Switzerland, this difference seems not only statistically significant but also economically relevant. This result is robust to the inclusion of firm characteristics, location characteristics, and variables mapping the local demand for apprenticeship positions. Moreover, we present various different specifications and subsample results, which do not alter our results.

Additionally, we examine various factors along the language border that potentially drive our results. First, we are concerned that our results stem from a discontinuity in the demand for apprenticeship positions also occurring at the language border. However, analyses of labor market outcomes, apprentice wages, and selection into the apprenticeship track yield no substantial differences among language regions and thus mitigate concerns to that effect. In contrast, we find evidence for different strengths of norm internalization between the language regions: German-speaking individuals show considerably higher financially unrewarded private commitment and prefer private- over state-owned businesses more than their Romance-speaking counterparts. Finally, we show that religious affiliation interacts with neither our norm measurement nor the firms' level of apprenticeship provision, and thus does not threaten our identification strategy.

Overall, the chapter shows how norm-guided behavior potentially limits the scope of purely financially motivated behavior by firms. In our opinion, one can draw two policy implications from this result. First, persistent norms might strengthen the sustainability of the Swiss VET system against potential shocks to firms' cost-benefit ratio. Second, behavior bound by norms might hinder the export of VET to other countries even after they set up institutional frameworks to foster VET.

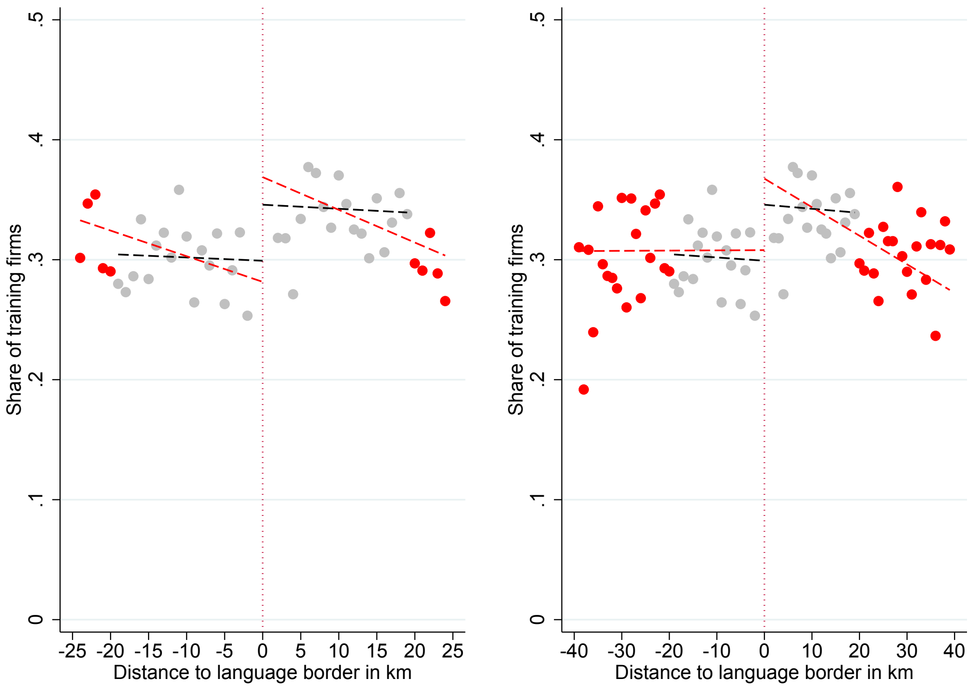
A Distance trend and bandwidth choice

Figure A.1: Share of training firms – RD sample and big cities



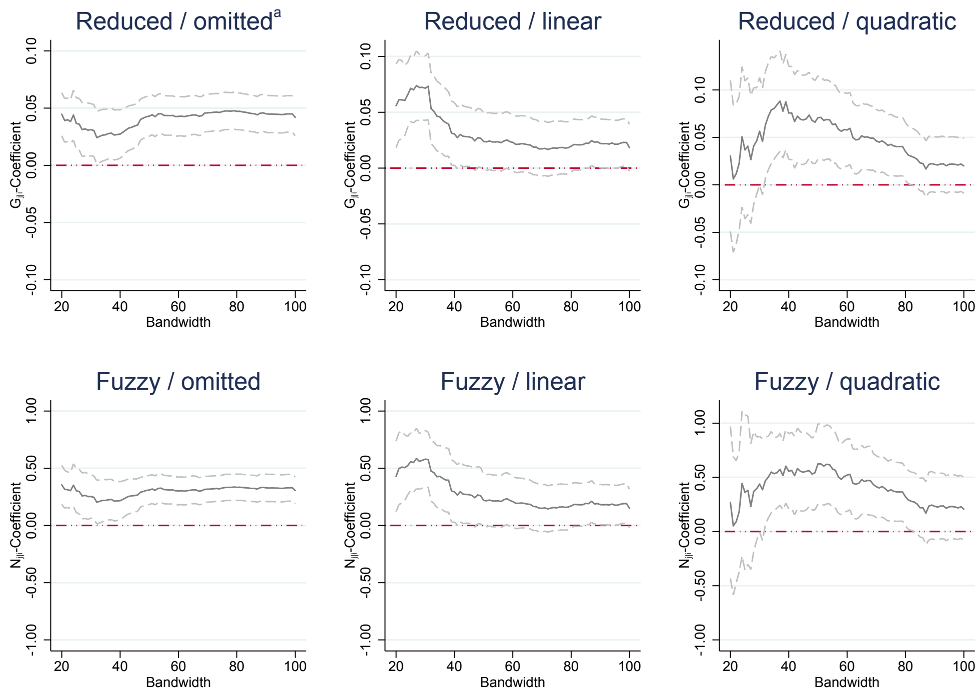
Notes: The figure shows the share of training firms aggregated, within bins of 1km width, by their distance to the language border in terms of actual travelling distance. German (Romance) speaking regions are associated with positive (negative) travelling distances. The black dashed horizontal lines show subsample means within 20 km bandwidth.

Figure A.2: Share of training firms – linear distance trends



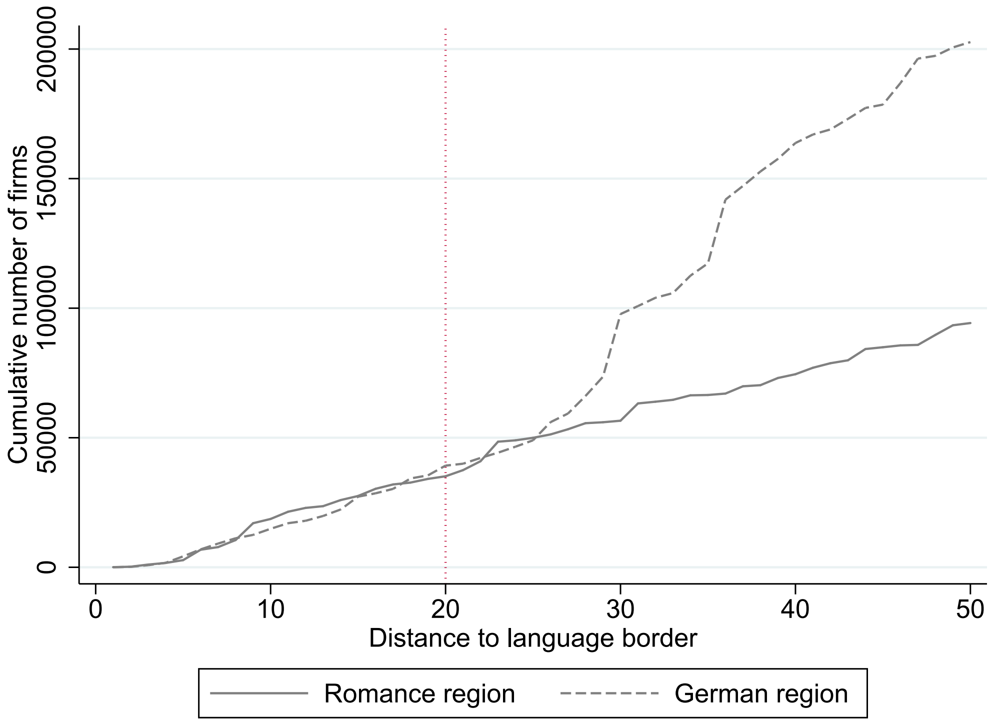
Notes: The figures show the share of training firms aggregated, within bins of 1km width, by their distance to the language border in terms of actual travelling distance. German (Romance) speaking regions are associated with positive (negative) travelling distances. The black (red) dashed lines show values from a linear regression within 20 (left figure: 25, right figure: 40) km bandwidth.

Figure A.3: Obtained coefficients over bandwidth



Notes: The figures in the upper (lower) row display the coefficients obtained from reduced-form (fuzzy) RD-estimations of the training dummy T_i on the German language dummy $G_{j[i]}$ (on the norm measurement $N_{j[i]}$ that is instrumented with the German language dummy $G_{j[i]}$). The reduced-form (fuzzy) estimations are specified as in model (6) of Table 2.4 (Table 2.5). [a] "omitted", "linear", and "quadratic" correspond to the distance to the language border measurement $d_{j[i]}$: The estimations yielding the coefficients displayed in column 1 omit $d_{j[i]}$, the estimations yielding the coefficients displayed in column 2 include $d_{j[i]}$, and the estimations yielding the coefficients displayed in column 3 include $d_{j[i]}$ and $d^2_{j[i]}$, respectively. Moreover, $d_{j[i]}$ and $d^2_{j[i]}$ are interacted with the German language dummy $G_{j[i]}$. For the corresponding fuzzy RD-estimates see also Table A.1. The grey dashed lines show the 95% confidence interval.

Figure A.4: Cumulative number of firms



Notes: The figure shows the cumulative number of firms, within bins of 1km width, by their absolute distance to the language border in terms of actual travelling distance.

Table A.1: Fuzzy RD-estimate over different bandwidths and with linear/quadratic distance trend (2SLS estimates)

	Distance trend omitted			Linear distance trend			Quadratic distance trend		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$N_{j[i]}$	0.355***	0.310***	0.306***	0.429***	0.219*	0.149*	0.269	0.624***	0.208
	(0.087)	(0.065)	(0.060)	(0.157)	(0.117)	(0.086)	(0.358)	(0.189)	(0.148)
$d_{j[i]}$				0.001	0.000	0.000	0.005	-0.002*	0.001
				(0.001)	(0.000)	(0.000)	(0.005)	(0.001)	(0.001)
$d_{j[i]} \times G_{j[i]}$				-0.002	0.000	0.001**	-0.007	0.000	-0.001**
				(0.002)	(0.000)	(0.000)	(0.008)	(0.002)	(0.001)
$d^2_{j[i]}$							-0.019	0.004*	-0.001
							(0.020)	(0.002)	(0.000)
$d^2_{j[i]} \times G_{j[i]}$							0.018	-0.008***	-0.001
							(0.036)	(0.003)	(0.001)
Constant	-0.662***	-0.055	0.026	-0.672***	0.010	0.160	-0.684***	-0.394**	0.066
	(0.183)	(0.123)	(0.107)	(0.190)	(0.135)	(0.117)	(0.196)	(0.182)	(0.161)
Bandwidth	20	50	100	20	50	100	20	50	100
Cantonal dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Census year dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Demand controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R^2	0.125	0.103	0.100	0.125	0.103	0.101	0.125	0.102	0.101
Observations	69,619	294,021	620,863	69,619	294,021	620,863	69,619	294,021	620,863
F-stat(First Stage)	124.0	294.9	339.1	64.45	104.2	181.9	21.03	57.89	64.14

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipality.

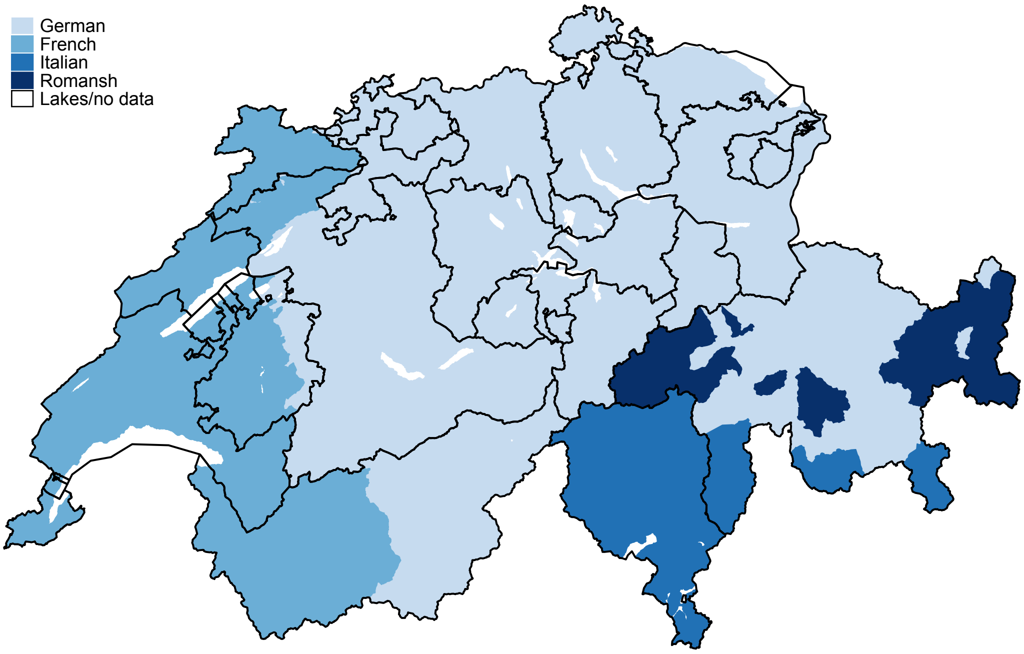
Table A.2: RD-estimates of all eight votes (OLS estimates)

	(1) Baseline	(2) Pub. health insurance	(3) Disability insurance	(4) Uni. health insurance	(5) Pub. postal service	(6) Old-age insurance	(7) Apprentice initiative	(8) Maternity coverage	(9) VET act
G_j	0.109***	0.142***	0.066***	0.140***	0.130***	0.111***	0.090***	0.160***	0.031
	(0.013)	(0.022)	(0.021)	(0.022)	(0.020)	(0.020)	(0.017)	(0.026)	(0.022)
d_j	0.001	0.002	0.000	0.001	0.000	0.001	0.000	0.001	0.002
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)
$d_j * G_j$	-0.002	-0.002	-0.000	-0.002	-0.002	-0.004**	-0.002	-0.000	-0.001
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Constant	0.534***	0.566***	0.600***	0.429***	0.268**	0.365***	0.627***	0.511***	0.908***
	(0.068)	(0.114)	(0.110)	(0.114)	(0.105)	(0.105)	(0.091)	(0.134)	(0.112)
Bandwidth	20	20	20	20	20	20	20	20	20
Cantonal dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Location characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dependent variable	N_j	*No-votes*	*No-votes*	*No-votes*	*No-votes*	*Yes-votes*	*No-votes*	*No-votes*	*No-votes*
R^2	0.715	0.713	0.492	0.730	0.547	0.470	0.466	0.709	0.238
Observations	379	379	379	379	379	379	379	379	379

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. Municipality level OLS-regression of share of no-votes (yes-votes in column 6) $VoteShare_j$ in respective referendum on German language region dummy G_j, the distance to the language border d_j, and their interaction term $d_j * G_j$. Source: Statistical Office 2017, own calculations.

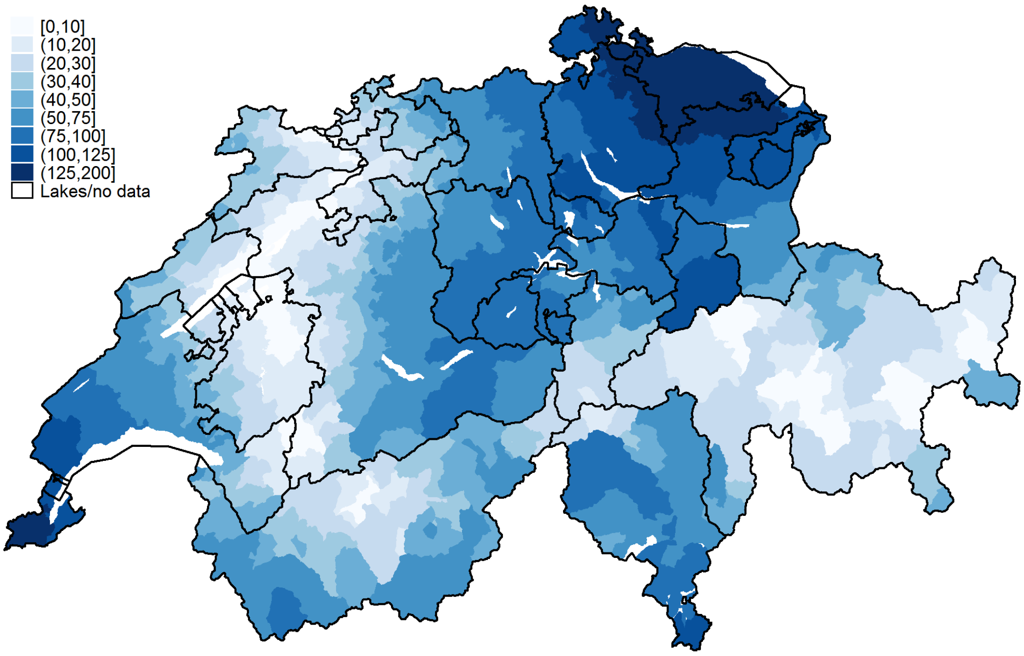
B Additional tables and figures

Figure B.1: Language regions in Switzerland



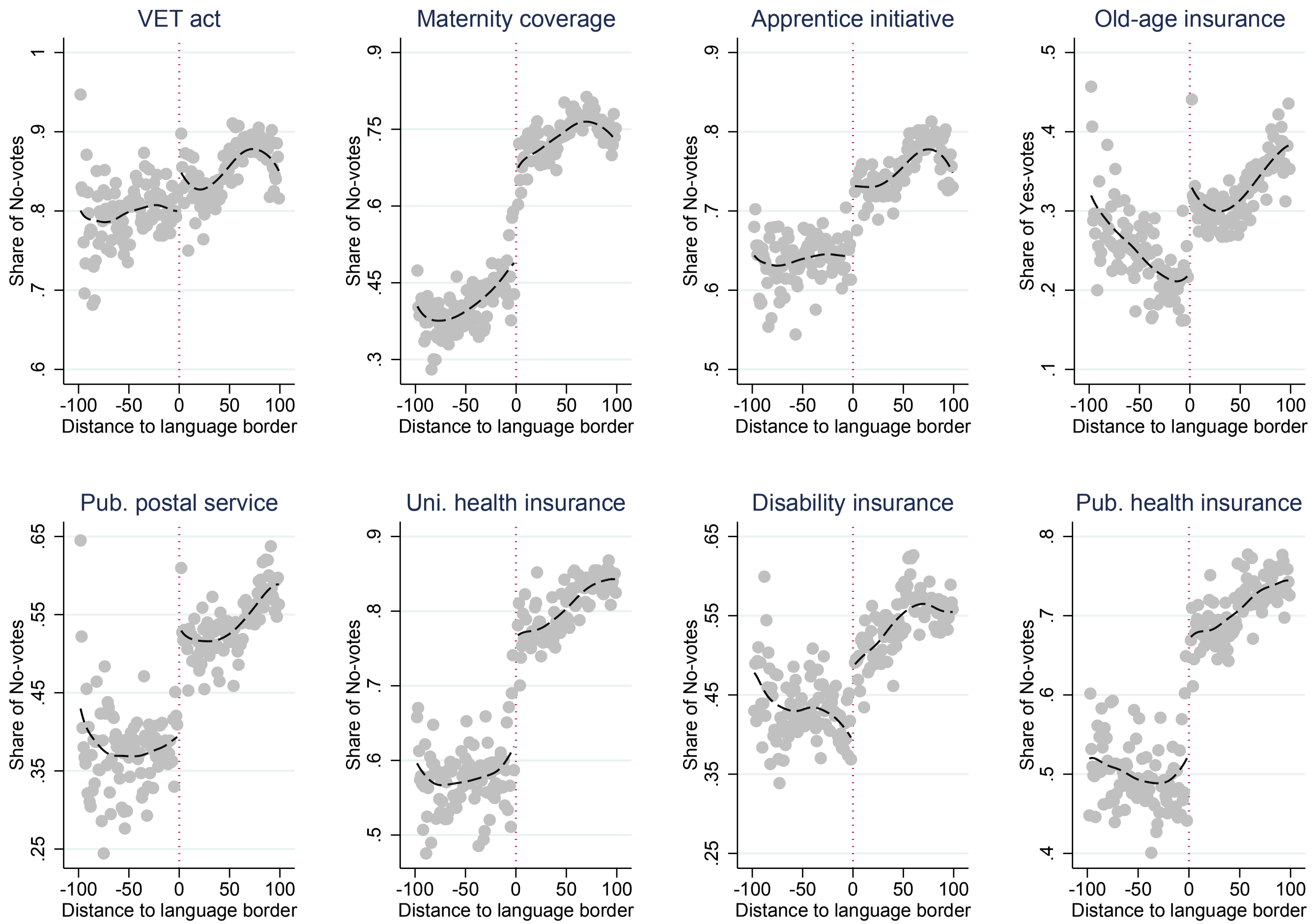
Notes: This figure maps the four different language regions (German, French, Italian, and Romansh) of Switzerland. The black lines delineate cantonal borders.

Figure B.2: Distance to language border in kilometres



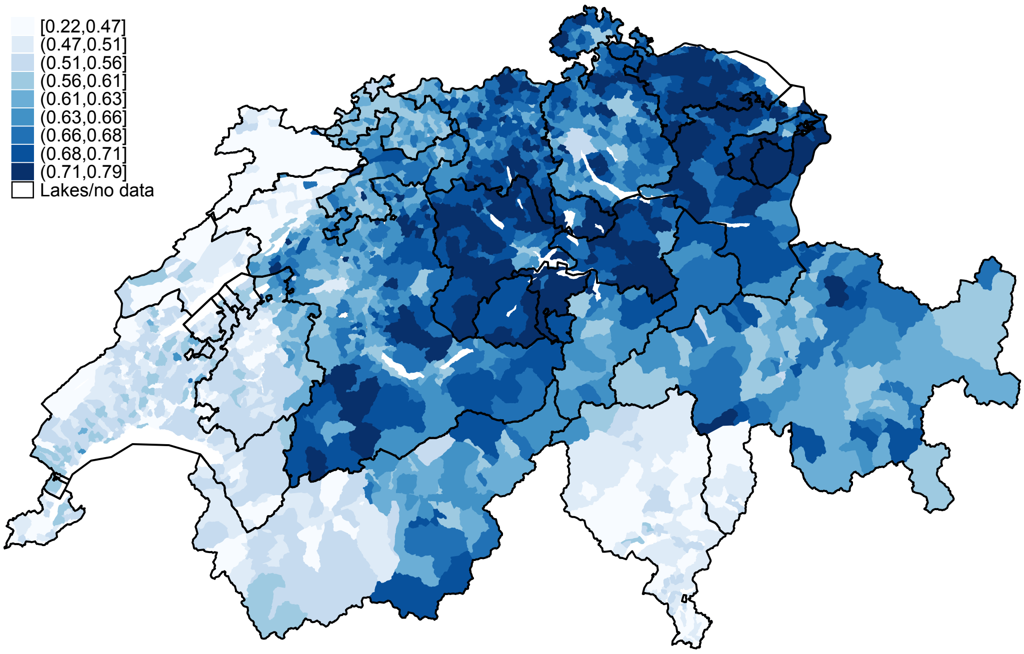
Notes: The figure shows the minimum travelling distance from a given municipality to the language border. Areas shaded in dark (light) blue are far away from (close to) the language border (cf. figure 2.1 in the main text). Black lines delineate cantonal borders.

Figure B.3: Discontinuities in voting results at the language border



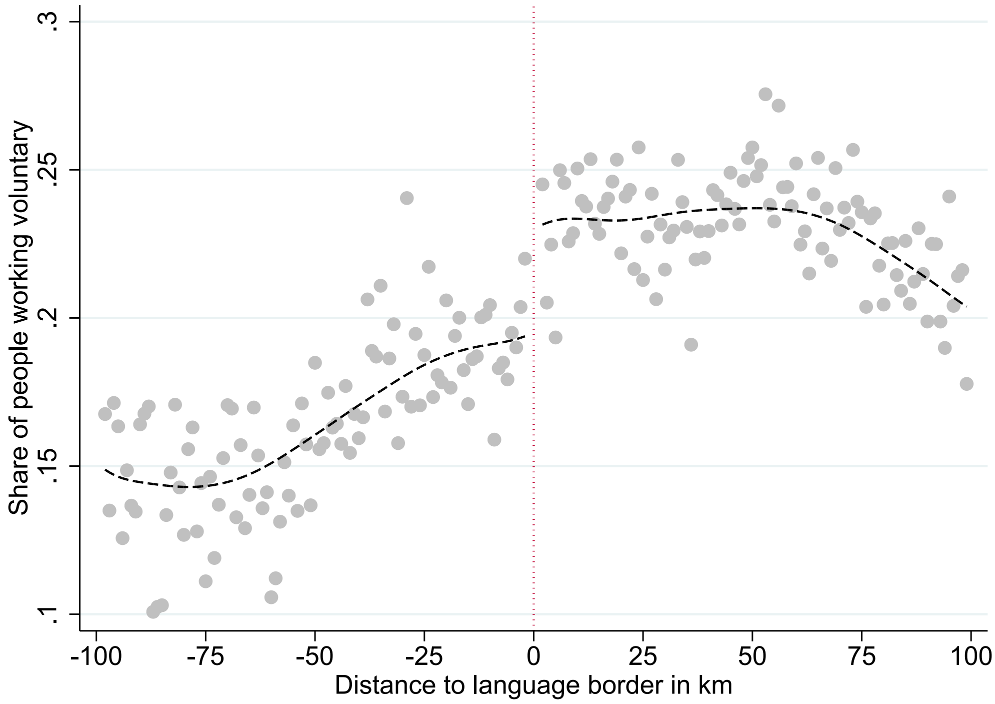
Note: The figure plots share of no-votes (yes votes in Old-age insurance vote) for municipalities aggregated, within bins of 1km width, by their distance to the language border in terms of actual travelling distance. German (Romance) speaking regions are associated with positive (negative) travelling distances. The dashed line shows smoothed values from a locally weighted regression.

Figure B.4: Norm measurement on the municipality level



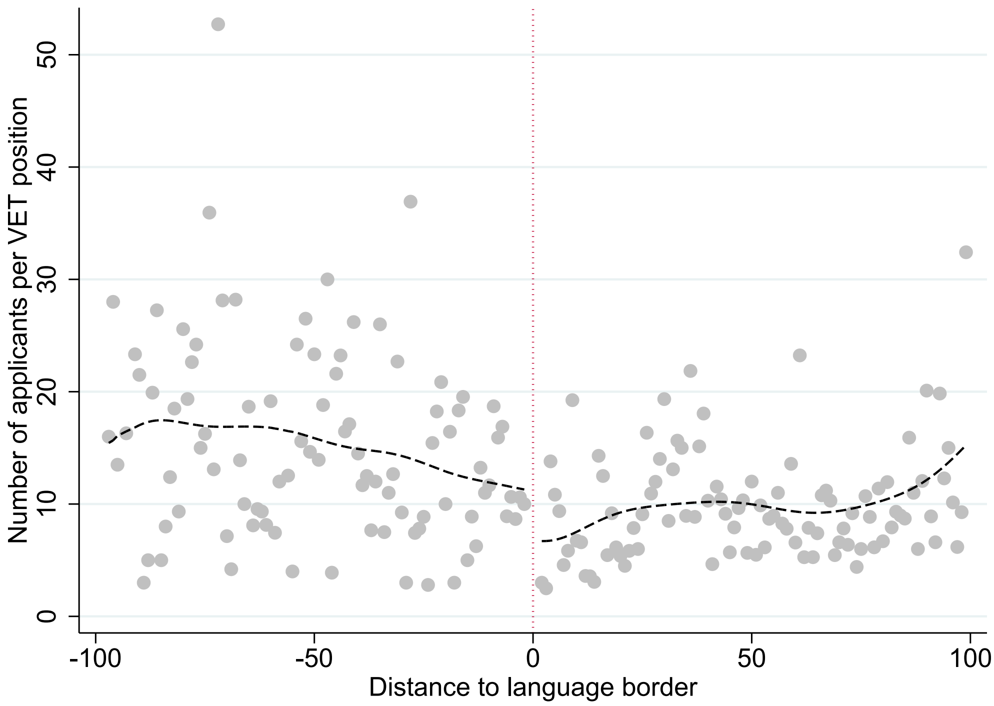
Notes: The figure shows the spatial distribution of $N_{j[i]}$ at the municipality level. Areas shaded in dark (light) blue correspond to regions with a strong (weak) norm towards private engagement. Black lines delineate cantonal borders.

Figure B.5: Discontinuity in the share of people working voluntarily at the language border



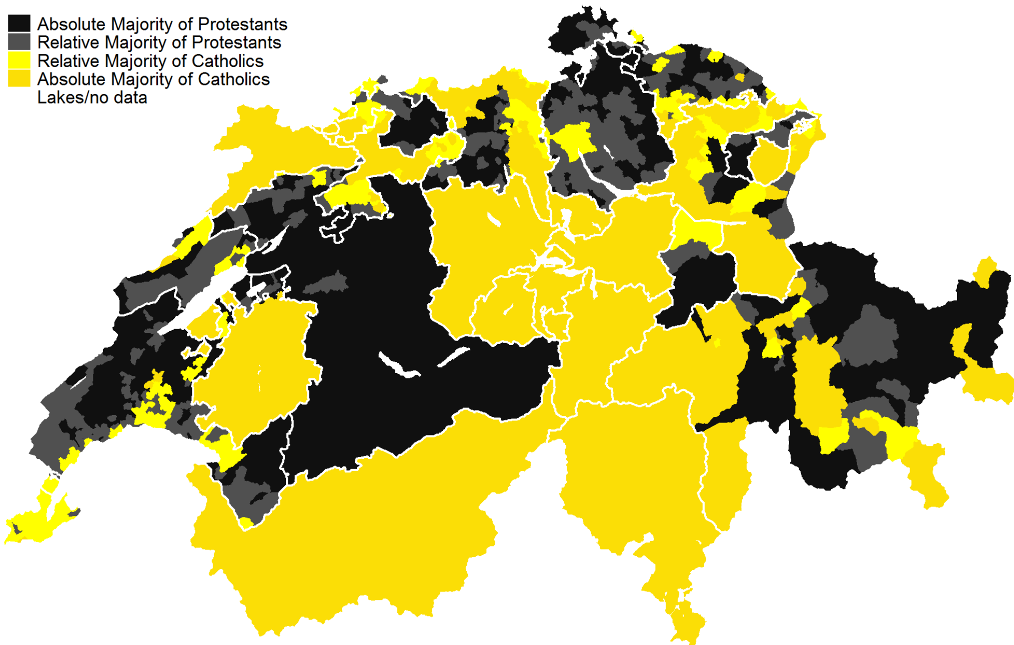
Notes: The figure shows the share of individuals working voluntarily, aggregated in bins of 1km width, by their distance to the language border in terms of travelling distance. German (Romance) speaking regions are associated with positive (negative) travelling distances. The dashed line shows smoothed values from a locally weighted regression.

Figure B.6: Discontinuity in the number of applications per vacant apprenticeship position



Notes: The figure shows the log number of applications per vacant apprenticeship positions, aggregated in bins of 1km width, by their distance to the language border in terms of actual travelling distance. German (Romance) speaking regions are associated with positive (negative) travelling distances. The dashed line shows smoothed values from a locally weighted regression.

Figure B.7: Protestantism and catholicism in Switzerland



Notes: The figure shows the religious affiliation at the municipality level in 2000. Black lines delineate cantonal borders.

Table B.1: Descriptive statistics (RD sample only)

	All	Romance	German	Difference[a]
(a) Main variables				
Training firm	0.322	0.302	0.342	0.040***
Norm towards private engagement	0.568	0.501	0.629	0.128***
(b) Firm characteristics				
Number of employees	13.7	14.1	13.3	-0.85
% For-profit firms	0.82	0.81	0.83	0.02***
% Public firms	0.13	0.14	0.12	-0.02***
(c) Location characteristics				
Log(number of firms)	7.60	7.74	7.47	-0.27
Log(same-sector firms)	5.85	6.01	5.71	-0.30
Log(Inhabitants)	7.05	7.08	7.01	-0.07
Population density	269.6	242.4	293.9	51.5
% Employed	0.61	0.61	0.61	0.01
% Employed in 2nd sector	0.34	0.34	0.34	-0.00
% Empolyed in 3rd sector	0.47	0.47	0.47	-0.00
Median income in CHF	6748.9	6619.6	6864.3	224.8
(d) Apprenticeship demand controls				
Distance high school	17.0	17.3	16.7	-0.60
Distance dual VET school	12.2	11.0	13.3	2.3***
Distance school-based VET	15.7	11.9	19.1	7.1***
% Age 15-25	0.11	0.10	0.11	-0.01
N (firms)	69,619	34,127	35,492	
N (municipalities)[b]	371	175	196	

Notes: The table shows mean values across the RD sample (bandwidth of 20 km around the language border). [a]Estimated differences of firm characteristics are clustered at firm level; estimated differences of location characteristics and apprentice supply controls are evaluated on municipality level and clustered at LM-region level if conducted on LM-region level. ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. [b]The number of municipalities does not correspond to the total number of municipalities because in this table of covariates we only consider municipalities with a least one firm and which thus enter the estimations.

Table B.2: List of votes used to measure the local norm favoring private engagement, $N_{j[i]}$

Nr.	Date	Title/description	Result	Share of supporting votes	Turnout	Measurement of $N_{j[i]}$
340	28.09.1986	Popular initiative for a "secured vocational education and training and retraining"	Rejected	18.4%	34.8%	No-votes
458	13.06.1999	Federal maternity insurance act	Rejected	39.0%	45.9%	No-votes
503	18.05.2003	Popular initiative "for a sufficient supply of vocational training"	Rejected	31.6%	39.6%	No-votes
507	16.05.2004	Federal law on old age insurance (11th AHV revision)	Rejected	32.1%	50.8%	Yes-votes
512	26.09.2004	Popular initiative "postal service for all"	Rejected	49.8%	53.5%	No-votes
528	11.03.2007	Popular initiative "for a unified social health insurance fund"	Rejected	28.8%	45.9%	No-votes
543	27.09.2009	Federal decree on temporary additional funding of disability insurance by raising VAT rates	Accepted	54.6%	41.0%	No-votes
586	28.09.2014	Popular initiative "for a public health insurance fund"	Rejected	38.2%	47.0%	No-votes

Notes: The vote number corresponds to the official numbering of the votes used by the Swiss Federal Administration. The share of supporting votes equals the fraction of all valid votes cast that were in favor of the vote, while turnout describes the fraction of eligible voters taking part in the vote. Additional information on the substantive issues involved in these votes that can be found online: urlhttps://www.bfs.admin.ch/bfs/en/home/statistics/politics/popular-votes.assetdetail.6666048.html.

Table B.3: Reduced form RD-estimates: evaluated at the "Jassgrenze" in the Canton of Aargau (OLS estimates)

	(1)	(2)	(3)	(4)
$C_{j[i]}$	-0.006	0.001	0.031	0.031
	(0.010)	(0.021)	(0.024)	(0.030)
$d_{j[i]}$			0.001	-0.000
			(0.002)	(0.002)
$d_{j[i]} \times C_{j[i]}$			0.003	0.004*
			(0.002)	(0.003)
Bandwidth	20	20	20	20
Census year dummies	No	Yes	No	Yes
Firm characteristics	No	Yes	No	Yes
Location characteristics	No	Yes	No	Yes
Demand controls	No	Yes	No	Yes
Observations	47991	47991	47991	47991
R^2	0.000	0.089	0.001	0.089

Notes: ***, **, and * denote statistical significance at the 1%, 5%, and 10% level, respectively. Robust standard errors are given in parentheses and are clustered by municipality.

Chapter 3

A task-based approach to horizontal mismatch

3.1 Introduction

Many economists claim that technological change has accelerated recently and transformed the labor market quite dramatically (Berger and Frey, 2016; Brynjolfsson and McAfee, 2012). Frey and Osborne (2017) estimate the probability of substitution for US occupations and postulate that 47% of the total US employment is at high risk. Using the same approach, Deloitte (2016) calculate similar numbers for Switzerland, Bonin *et al.* (2015) for Germany. However, Frey and Osborne's (2017) approach might focus too narrowly on the substitution effects of new technology and underestimate its complementarity to human capital. Autor (2015) and Acemoglu and Restrepo (2018) describe how this complementarity between automation and labor increases productivity, raises earnings, and augments demand for labor. Accordingly, they argue that new technologies do not lead to fewer jobs but alter their nature (see also Bessen, 2016; Evangelista *et al.*, 2014; Graetz and Michaels, 2015; Gregory *et al.*, 2016; Hirsch-Kreinsen, 2016). In a labor market where the importance of certain occupations erodes and the characteristics of others change rapidly, individuals working in an occupation different from that they learned, termed occupational horizontal mismatches, become more likely.

In this chapter, I analyze first how recent technological change affects individuals' horizontal mismatch probabilities and second the extent to which these horizontal mismatches translate into wage penalties. In order to trace this mechanism and to identify its causal effect on wages, I propose an instrumental variable (IV) approach. The task composition of individuals' learned occupation serves as the instrument for the endogenous mismatch variable in this case. This empirical strategy leans on the task-based approach developed by Autor *et al.* (2003), which shows that technology increases and decreases demand for complementary and substitutable tasks and occupations bundling them, respectively.

Applying this strategy to a sample of more than 10,000 observations of roughly 1,200 Swiss males in the years 1999-2016 reveals a mismatch wage penalty of roughly 12%. This estimated mismatch wage penalty pertains to individuals who primarily learned substitutable occupations and are thus negatively affected by task shifting technological change. In contrast, I find no mismatch wage penalties for individuals who learned mostly complementary or unaffected occupations. Comparisons of different types of mismatches and different strategies (ordinary least square OLS, fixed-effect, and IV) to identify them, presented in this chapter, suggest that these distinctions are relevant for any conclusion on mismatch wage penalties.

This chapter relates to the existing mismatch literature, which can broadly be divided into two strands (for an overview see Somers *et al.* (2019)). One strand relies merely on OLS estimations and generally associates horizontal mismatches with a wage penalty. Various personal characteristics and the strength of the mismatch thus determine the magnitude of this association. Robst (2007b) estimates wage penalties of roughly 11% for employees with formal qualifications that are "not related" to their occupation and roughly 2% for employees with formal qualifications that are "somewhat related". Similarly, Bender and Heywood (2009), Nordin *et al.* (2010), Bender and Roche (2013), and Yakusheva (2010) find wage penalties that increase with individuals' mismatch intensity. Zhu (2014) explains the small wage penalty of 1.2% for Chinese males and 1.5% for Chinese females with an educational system that provides graduates with mostly general skills. Similarly, Nordin *et al.* (2010) show that employees holding degrees providing mostly job-specific skills suffer from the largest wage penalties when experiencing horizontal mismatch. Bender and Heywood (2009) highlight how the specificity of acquired human capital increases potential wage penalties at later stages of PhDs' careers. Meanwhile, wage penalties seem to disappear over time as mismatched individuals acquire occupation-specific human capital within their new occupation (Malamud, 2010).

Another strand of the mismatch literature applies fixed-effect estimations to account for unobservable personal characteristics, e.g. ability. Schweri *et al.* (2019) estimate no mismatch wage penalty for Swiss males either considering themselves mismatched or being objectively mismatched (different learned and current occupations). Bender and Heywood (2009) apply fixed-effect estimations to a sample of US doctoral graduates and find a small wage penalty for males and a small but statistically insignificant wage penalty for females. However, I argue that these fixed-effect estimations suffer from one main shortcoming: while some individuals become mismatched due to lay-offs, others might choose voluntarily to become mismatched, for instance because they concurrently realize a wage gain. Accordingly, Robst (2007a) attributes an overall negative effect for males of 10.2% and an overall negative effect for females of 8.9% to different types of mismatches: male employees experiencing a mismatch because "no matching job is available" suffer from a wage loss of 26.5% (female employees: 18.5%), and employees experiencing a mismatch because of the "job location" earn 29.3% (21.1%) less than their matched counterparts. Contrarily, male employees becoming mismatched because of payment or promotion opportunities realize a wage gain of 6.1% (9.1%).

The contribution of this chapter to this existing mismatch literature is twofold. First, I propose a mismatch measurement that goes beyond a mere binary mismatch measurement. Instead, I use detailed survey data on occupational skill portfolios to estimate the strength of horizontal mismatches between any pair of learned and current occupations. Thus, I allow for wage penalties that vary with the magnitude of human capital loss induced by horizontal mismatches. Applying such a continuous horizontal mismatch measurement is motivated by the work of Guvenen *et al.* (2015), Bender and Heywood (2009), Nordin *et al.* (2010), Bender and Roche (2013), Robst (2007b), and Yakusheva (2010), who argue that employees who can partly make use of their acquired skills despite being horizontally mismatched must only accept a small wage penalty. By comparing my results with estimated wage penalties based on a mismatch dummy, I show that understanding horizontal mismatch as a binary concept might overstate mismatch wage penalties. Moreover, the continuous mismatch measurement partly mitigates concerns arising from potential measurement error likely inherent in any occupational coding in survey data. Second, I propose an IV approach to estimate a wage penalty stemming causally from horizontal mismatches. To the best of my knowledge, all attempts to estimate *causal* mismatch wage penalties rely on fixed-effect estimations. However, Section 3.4 of this chapter demonstrates theoretically how fixed-effect estimations merely reveal an effect on average that potentially underestimates wage effects for involuntarily mismatched individuals. Comparisons of IV estimations and fixed-effect estimations applied to the same sample in Section 3.5 underpin this empirically.

The remainder of this chapter is organized as follows. Section 3.2 introduces the task-based approach and relates it to the concept of horizontal mismatch. Section 3.3 describes the three data sources used in this chapter. Section 3.4 outlines the empirical strategy. Section 3.5 presents the estimation results, and Section 3.6 concludes.

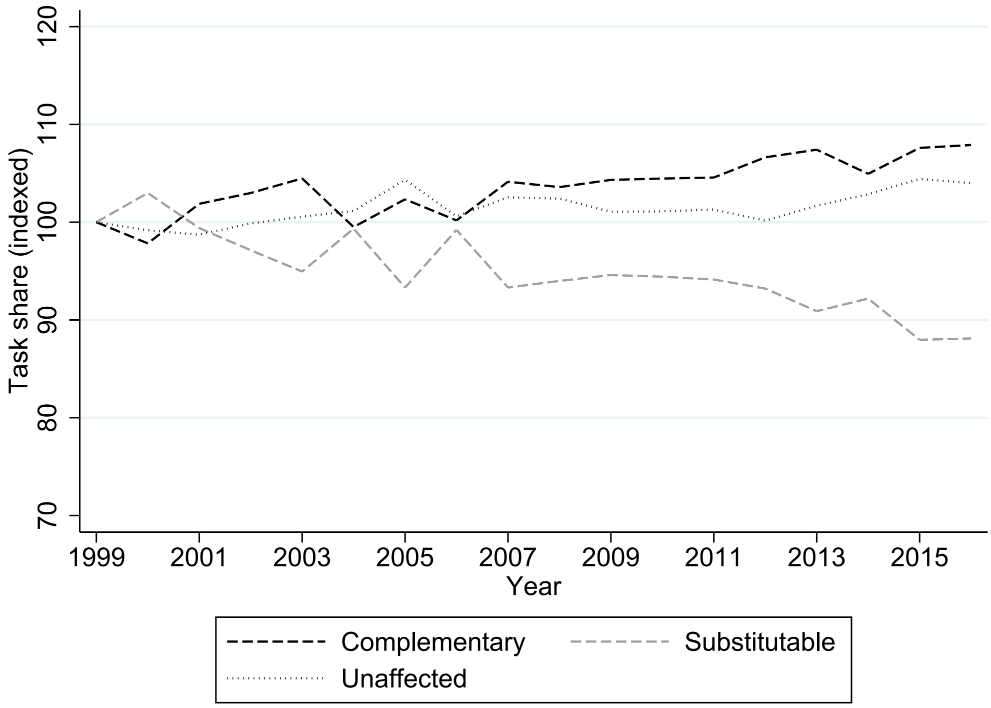
3.2 Theoretical background

The task-based approach was introduced by Autor *et al.* (2003) "to study how computerization alters job skill demands (p.1279)." They argue that "present computer technology is more substitutable for workers in carrying out routine tasks than non-routine tasks, it is a relative complement to workers in carrying out non-routine tasks" (p. 1285). In light of falling computer prices, both these mechanisms increased the relative demand for workers executing nonroutine tasks, typically college graduates. This was especially striking because most previous literature described an increasing college premium (for an overview see Katz, 1999) but failed to provide insights on technology-related mechanisms enhancing it.

Beside substitutable "cognitive and manual routine tasks" and complementary "analytical and interactive non-routine tasks", Autor *et al.* (2003) introduced a fifth task category: manual nonroutine tasks that are little affected by new technology. Either these tasks are not substitutable because they are too complex to be executed by machines (e.g. cleaning of different room types), or technology is barely complementary to the execution of these tasks because they require interpersonal communication that neither machines nor computers can deliver (e.g. psychological consultation). This general pattern of a strong increase in the demand for nonroutine analytical and interactive tasks, a stable or slightly increasing demand for nonroutine manual tasks, and a shrinking demand for cognitive and manual routine tasks was first described for the US by Autor *et al.* (2003), for Germany by Spitz-Oener (2006), and for the UK by Goos and Manning (2007); for an overview see Autor (2013).

In this chapter, I strongly lean on the task categorization introduced by Autor *et al.* (2003) but distinguish only three task types. I consider tasks in which technology supports people who carry them out to be *complementary*. This complementarity between humans and technology increases the efficiency of these tasks' execution and thus the return for individuals performing them. In Autor *et al.*'s (2003) categorization, these tasks are referred to as “analytical nonroutine” and partly as “interactive nonroutine” tasks. *Substitutable* tasks are executable by computers or machines without or with little help from humans and hardly restricted by financial, legal, or ethical constraints. Humans performing these tasks tend to be replaced by machines, and their expected return decreases. The substitutability of these tasks is what classifies them as “cognitive routine” or “manual routine” tasks in Autor *et al.*’s (2003) categorization. Finally, technology plays little or no role in the execution of *unaffected* tasks. This little impact accounts for both the demand for and the return on these tasks. Autor *et al.*’s (2003) categorizes these tasks as “manual nonroutine” and partly “interactive nonroutine” tasks. Such a three-type categorization, which is somewhat similar to that proposed above, can be found in Autor and Handel (2013).

Figure 3.1: Task shares in Switzerland 1999 - 2016



Notes: The figure shows the indexed shares of complementary, substitutable, and unaffected tasks over the sample period. The calculation of the task shares is explained in Section 3.3.3.

To categorize these three tasks types, I exploit the Swiss Job Market Monitor, which contains a representative sample of job vacancies and lists the most important task requirements for every vacant position; for details see Section 3.3.3. Figure 3.1 displays the development of the three task types described above in the Swiss labor market between 1999 and 2016. As expected, tasks complementary to new technology increased their share by 8.8%, while tasks substitutable by new technology decreased their share in the labor market by 13.4%. The share of tasks unaffected by new technology increased slightly (+4.6%). This pattern fits the findings in the literature discussed above and previous findings on task shifts in the Swiss labor market (Aepli *et al.*, 2017). As for example Manning (2004) (for Switzerland: Oesch and Rodriguez Menés (2010) and Aepli *et al.* (2017)) do, I argue that technological change affecting firms' production processes at least partly triggers these task shifts.[1]

Most task-based literature exploits these shifts in task demand to explain how technology either affects the employment structure (e.g. Acemoglu and Autor, 2011; Autor *et al.*, 2003; Autor, David and Dorn, 2013; Goos and Manning, 2007; Gregory *et al.*, 2016; Spitz-Oener, 2006) or shifts the earning distribution due to varying task returns (e.g. Acemoglu and Autor, 2011; Dustmann *et al.*, 2009; Firpo *et al.*, 2011). The present chapter applies the task-based approach to an intersection of these two labor market

[1]However, technological change, understood in a narrow sense, might not be the only driver for these task shifts in labor markets across highly developed economies. For example, Goos *et al.* (2014), Harrison and McMillan (2011), and Pierce and Schott (2016) stress the importance of reallocation of industrial production to China or Eastern-European countries for the disappearance of relatively simple, manual routine tasks in developed countries; for Switzerland see Waser and Hanisch (2011). However, the reallocation of industrial production to other countries also often requires some sort of new technology, such as transportation, ICT, or transfer of relatively new technology to these countries. Speaking of technological change thus includes outsourcing to other countries to at least some extent. Therefore, a more accurate term could be task-shifting technological change. Moreover, demand factors are likely to play a role. For example, ageing societies increase the need for care professions (Degen and Hauri, 2017), which largely consist of manual nonroutine tasks that are unaffected.

developments: horizontal mismatches. In the present chapter, I understand horizontal mismatches as the situations of individuals whose educational field or learned occupation does not match their current occupation (for an overview see Somers *et al.*, 2019). More generally speaking, this refers to a mismatch between individuals' acquired skills (supply-side), their human capital, and the tasks demanded in the labor market (demand-side). These mismatches are presumably associated with wage losses and thus affect both the occupation people work in and how much they earn for their labor.

Technology affects these horizontal mismatches in two ways. First, technology is considered to be the main demand-side driver for task shifts (Manning, 2004), whereas international trade and product demand shifts play a minor role (OECD, 2005). Thus, given the supplied skills, task-shifting technology change affects the match between the demanded tasks and the supplied skills. In the mismatch literature, Bender and Heywood (2011) report high mismatch probabilities for engineers and hard scientists, which they attribute to rapid technological change within these fields. Similarly, Witte and Kalleberg (1995) argue that skills acquired during an apprenticeship that become obsolete due to changing task requirements foster occupational mismatches. Second, when technology alters returns to tasks, it triggers occupational mobility, so that individuals select into tasks displaying increasing returns (e.g. Autor and Handel, 2013). In general, this arguably increases the overall match quality, as task prices enhance their efficient allocation. However, high-ability workers selecting into well-paid positions might in turn push other workers out of their traditional occupations and thus trigger mismatches among them.

Section 3.4.4 describes specifically how task shifts are related to horizontal mismatches. Section 3.5.1, in what is the first-stage in my IV approach, estimates the direction and magnitude of this relation.

3.3 Data

The empirical analysis in this chapter is based on three data sources. The Swiss Household Panel (SHP) observes the population of interest and is described in Section 3.3.1. I derive a continuous mismatch measurement between any occupation pair from the German BIBB/BAuA Employment Survey (BIBB/BAuA-ES). Section 3.3.2 presents this data source and highlights the benefits of such a continuous measurement. To determine the occupational exposure to technology, I rely on the Swiss Job Market Monitor introduced in Section 3.3.3.

3.3.1 Swiss Household Panel

The Swiss Household Panel (SHP) is the main data source used in this chapter; it surveys a representative sample of Swiss households between 1999 and 2016. Although respondents were asked about other household members in some domains, I only rely on the actual respondents' information for the analysis below, and thus my observation units are individuals. Beside detailed demographic information, the SHP covers various information on individuals' labor market status, such as individuals' wage, education, firm tenure, and hierarchical level (see Table 3.1). These variables allow me to estimate a basic Mincer equation that forms the basis of the econometric strategy introduced in Section 3.4. Moreover, the SHP collects a set of employer characteristics, including the overall number of employees and the firm's industry. To construct my main independent variable, the mismatch variable, I rely on a subsample of individuals for whom the SHP includes retrospective information on education and work episodes. This biographical subsample contains 28,469 observations stemming from 3,249 working individuals.

Due to women's selective participation in the labor market, I restrict the sample to males between the ages of 20 and 65 with either a VET, a tertiary-B (further education for individuals with a VET degree), or a tertiary-A (university or university of applied science) degree. Additionally, I exclude

Table 3.1: Descriptive statistics

	Full sample	Matched	Mismatched	Difference
Individual characteristics				
Age	45.3	44.4	46.0	-1.63***
Have children	0.48	0.52	0.45	0.06**
Married	0.74	0.75	0.73	0.03
Foreign	0.09	0.09	0.09	0.00
VET	0.38	0.39	0.38	0.01
Tertiary-B	0.35	0.32	0.38	-0.06**
Tertiary-A	0.26	0.29	0.24	0.06**
Director	0.14	0.13	0.15	-0.02
Supervisor	0.69	0.70	0.69	0.01
Further Educ.	0.45	0.47	0.43	0.04**
Employment in %	96.5	96.4	96.6	-0.22
Firm informations				
<10 Emp.	0.12	0.12	0.12	0.01
10 - 49 Emp.	0.21	0.22	0.19	0.03*
50 - 99 Emp.	0.10	0.10	0.10	-0.00
100+ Emp.	0.48	0.46	0.49	-0.03
Industrial sector	0.27	0.28	0.26	0.01
Service sector	0.66	0.64	0.67	-0.03
Wage and mismatch				
Log monthly wage	8.97	8.94	8.99	-0.05**
Mismatch dummy	0.52	0.00	1.00	-1.00***
Occ. distance	0.53	0.00	1.00	-1.00***
Task shares learned occupation				
Complementary	0.37	0.38	0.36	-0.01
Substitutable	0.42	0.40	0.44	-0.04*
Unaffected	0.21	0.22	0.20	0.02
Task shares current occupation				
Complementary	0.44	0.40	0.47	-0.07***
Substitutable	0.30	0.36	0.25	0.10***
Unaffected	0.26	0.24	0.27	-0.03**
N(person-year)	10,471	4,988	5,483	-
N(person)[a]	1,224	457	525	-

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Most variables contain some missing values which are coded as such but not displayed in the table. This is why not all displayed shares sum up to 1. [a]Person observations are considered matched (mismatched) if a person is matched (mismatched) in every year she/he is observed.
Sources: SHP 1999 - 2016, SJMM 1999 - 2016, and BIBB/BAuA 2006/2012.

410 individuals with a workload below 50%, and 468 individuals with an annual income below 24,000 or above 300,000 Swiss Francs, respectively. Overall, this lowers my sample to 10,471 observation stemming from 1,224 individuals. For roughly half of these person-year observations, the learned and the current occupation are not the same, i.e. they are mismatched, as Table 3.1 highlights.[2]

3.3.2 Occupational distance measurement

Lazear (2009) argued in a seminal paper that all skills are basically general, but firms combine ("weight") them differently in their production processes. This varying skill demand across firms enhances the specificity of skills and thus of human capital; the more specific an individual's skill combination, the more specific his or her human capital. Geel *et al.* (2011) apply Lazear's (2009) skill-weight approach to three waves of the German BIBB/BAuA-ES dataset and determine occupations' specificity. The BIBB/BAuA-ES interviews employees, among other topics, about how intensely they perform various tasks at their workplace. Geel *et al.* (2011) argue that these task items approximate the skill portfolio that workers need to perform their jobs and, when aggregated within occupations, the skill portfolio of an occupation. In the empirical part of their paper, they then show how German graduates who learned a rather specific VET occupation shy away from leaving this occupation, presumably because the wage loss is expected to be large due to their specific skill combination.

In the vein of Geel *et al.* (2011), I argue in this chapter that the skill composition of two occupations determines the strength of a potential mismatch between them. This seems intuitive: the closer the skill combination of two occupations is, the smaller the wage penalty for an individual who works in one of these occupations but has learned the other one. Accounting for the potential heterogeneity of mismatches is not new in the mismatch literature. Guvenen *et al.* (2015), Bender and Heywood (2009), Nordin *et al.* (2010),

[2]Additionally, Figure A.2 in the Appendix displays the temporal evolution of mismatches during the analysis period.

Bender and Roche (2013), Robst (2007b), and Yakusheva (2010) all show that the mismatch wage penalty increases with the strength of the perceived mismatch.[3]

Besides accounting for the potential heterogeneity of mismatches, such a continuous mismatch measurement is beneficial for another reason. Occupational coding in any survey likely contains some measurement error (Bound *et al.*, 2001, p.3802); for example different interviewers might assign similar information to different occupational codes. In case this measurement error is random, it biases the estimated wage penalty estimate towards zero (Angrist and Pischke, 2014; Bound *et al.*, 2001). Though any continuous mismatch measurement relying on learned and current occupation codes also suffers from this measurement error, it seems plausible that wrongly coded individuals are misclassified into occupations close to their true occupation. Measurement error in occupational classification appears thus less severe when relying on a continuous rather than a binary mismatch measurement that weights all mismatches equally.

For the concrete construction of my continuous mismatch measurement, I rely on the same data as Geel *et al.* (2011) but employ the latest waves of 2006 and 2012,[4] and proceed as follows: (i) For each of 16 task items (see Table 3.2), I aggregate the answer ("How often does this task occur during your work?" 0=never; 0.5=seldom; 1=often) of 24,975 individuals[5] in the BIBB/BAuA-ES at the level of the 144 observed 3-digit German occupations. (ii) Within every occupation, I divide each task item value by the sum of all 16 task item values. Thus, every German occupation consists of 16 task item shares that sum up to one. (iii) Based on occupational

[3] In contrast, applying a binary mismatch concept, i.e. indicating mismatch with a dummy D_{it} that takes the value 1 if a person's current occupation does not correspond to any occupation this person has learned and 0 otherwise, cannot account for the heterogeneity of mismatches. See also the discussion in Section 3.4.2.

[4] https://www.bibb.de/veroeffentlichungen/de/publication/show/7094 and https://www.bibb.de/veroeffentlichungen/en/publication/show/2274.

[5] I exclude East Germany, observations with wages in the bottom or top 1 percentile, individuals working less than five hours per week, and observations with no occupational information.

Table 3.2: Task items in the BIBB/BAuA 2006/2012 surveys

	SHP sample statistics			
Items as in BIBB/BAuA 2006/2012	Mean	Standard deviation	Min.	Max.
Manufacture, production of goods and merchandise	0.313	0.255	0.015	0.875
Measuring, testing, quality control	0.633	0.170	0.264	0.909
Monitoring, control of machines, plants, processes	0.411	0.221	0.085	0.889
Repair, overhaul	0.358	0.180	0.046	0.826
Retail, procurement, selling	0.311	0.136	0.090	0.757
Transport, storage, dispatch	0.374	0.140	0.136	0.861
Advertising, marketing, public relations	0.260	0.148	0.057	0.772
Organizing, planning and preparing work processes	0.548	0.109	0.246	0.826
Development, research, construction	0.287	0.136	0.050	0.766
Training, teaching, educating	0.424	0.189	0.100	0.986
Collecting information, researching, documenting	0.657	0.183	0.219	0.933
Advising and informing	0.102	0.134	0.013	0.687
Hosting, accommodating, preparing food	0.186	0.200	0.017	0.927
Nursing, caring, healing	0.295	0.136	0.013	0.735
Securing, protecting, guarding, monitoring	0.743	0.196	0.183	0.999
Cleaning, waste disposal, recycling	0.380	0.191	0.045	0.819
N(Occupations)	86			

Notes: Task items values refer to 0=never, 0.5=seldom, and 1=often and are aggregated at the 3-digit level of 144 German occupations. These values are then converted to the 86 Swiss occupations (SSCO 3-digit level) observed in the SHP.

frequencies, I assign these task-item shares to a maximum of five learned occupations and to the current occupation of each individual at the 3-digit level of the Swiss Standard Classification of Occupations 2000 (SSCO 2000).[6] (iv) The occupational distance between any learned occupation

[6]Concerning the transferability of task items from German to Swiss occupations, I argue that, due to similar economic structure and a similar education system (e.g. the importance of the vocational track), German occupations resemble Swiss occupations more than, for example, US-occupations. Consequently, Marsden (1999) argues in his *theory*

($locc = 1, 2, 3, 4, 5$) and any current occupation ($cocc$) equals the sum of their absolute differences across each of the 16 occupational task-item shares, formally for individual i's learned and current occupation: $OccDist_{i,locc} = \sum_{j=1}^{16} |item_{i,j}^{locc} - item_{i,j}^{cocc}|$. (v) The occupational distance relevant to the mismatch analysis is the smallest difference between any of individual i's learned occupations at time t and individual i's current occupation at time t, formally: $OccDist_{it} = min\{OccDist_{it,locc=1}, ..., OccDist_{it,locc=5}\}$. (vi) Finally, I normalize this continuous mismatch measurement over the sample to mean one for all individuals considered mismatched, i.e. for $OccDist_{it} > 0$. Thus, any estimated wage penalty based on this continuous mismatch measurement can be interpreted as switching from a match to an average mismatch in terms of occupational distance $OccDist_{it}$. Figure 3.2 displays the distribution of this occupational distance for the mismatched subpopulation, i.e. with $OccDist_{it} > 0$.

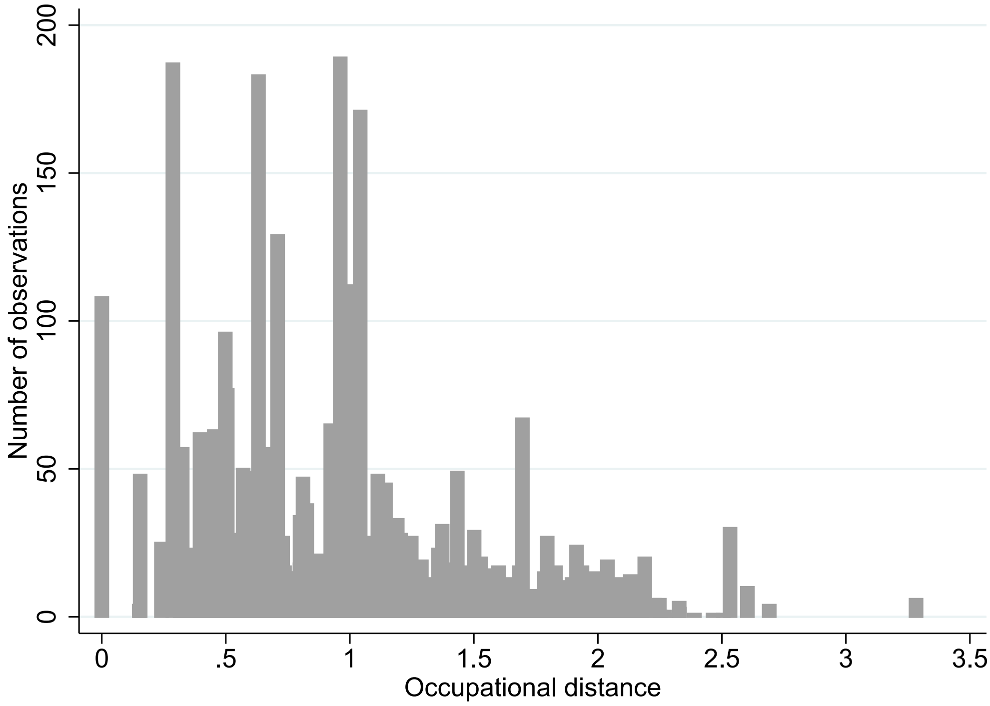
For comparison purposes, I primarily display estimates based on a conventional mismatch dummy, D_{it}. However, I suggest focusing on estimations applying the occupational distance as a continuous measurement for horizontal mismatch.

3.3.3 Occupational task shares

In order to determine the occupational task composition in the Swiss labor market, I exploit the Swiss Job Market Monitor (SJMM), conducted annually by the University of Zurich. The SJMM is a representative monitor of vacant positions published by firms on online job portals, in newspapers, and on their websites. Beside general job requirements and characteristics, the SJMM lists the task considered to be the most relevant for a vacant position out of a set of 21 possible tasks. To the best of my knowledge, this is the only information

of employment systems that the German labor market, which is based on occupational qualifications, allows a high interfirm mobility of skilled workers and adjusts fast to technological change. In contrast, the labor force in the US labor market receives more on-the-job training and stricter guidelines. According to Marsden (1999), the Swiss labor market belongs to the same category as the German labor market. The scheme for recoding German occupations to Swiss ones is available upon request.

Figure 3.2: Distribution of the continuous mismatch measurement



Notes: The figure shows a histogram of the occupational distance measurement over the subsample of mismatched individuals (5,483 person-year observations). Person-year observation of matched individuals excluded (n=4,988). The calculation of the occupational distance measurement is explained in Section 3.3.2.

available on the task content of Swiss occupations. The sample I use consists of 43,932 observations collected during 1995 and 2015. 42.4% of these job vacancies were published on firm websites, 36.5% in newspapers, and 21.1% on job portals.

The procedure of building broader task categories based on surveys that collect individual task data is called a survey-based[7] task-based approach, and was first introduced by Spitz-Oener (2006); moreover, it can be found in Autor and Handel (2013). Both these works are based on the conceptual

[7]Other task-based approaches rely on experts' assignments of tasks based on descriptions or curricula of occupations, e.g. Acemoglu and Autor (2011), Autor *et al.* (2006), and Goos and Manning (2007). For a discussion on the two different approaches see for example Autor and Handel (2013) and Rohrbach-Schmidt and Tiemann (2013).

Table 3.3: Task items in the SJMM and their categorization

Items	Mean[a]	Standard deviation	Min.	Max.	Experts' task categorization[a]
Cultivation, breeding, wining/dismantling	0.043	0.170	0	0.867	2 / 2 / 2
Handicraft/machine production	0.127	0.219	0	1	2 / 2 / 2
Install, assemble, build	0.050	0.135	0	0.769	2 / 2 / 2
Setup, programming, control, operation	0.081	0.164	0	0.730	3 / 2 / 2
Repair, maintain, restore	0.030	0.088	0	0.725	3 / 2 / 2
Warehousing, shipping, transport	0.045	0.152	0	0.930	3 / 2 / 2
Buy/sell, collect (cash), advise customers	0.091	0.184	0	0.883	3 / 2 / 3
Writing, correspondence, edit forms	0.037	0.086	0	0.538	1 / 3 / 3
Calculate, keep accounts	0.012	0.044	0	0.332	2 / 3 / 2
EDP activities, programming	0.014	0.074	0	0.652	1 / 1 / 1
Serve, host	0.015	0.064	0	0.533	3 / 3 / 3
Ironing, cleaning, waste disposal	0.025	0.110	0	0.779	3 / 3 / 3
Secure, guard	0.012	0.079	0	0.724	3 / 2 / 3
Analyse/research, review	0.073	0.158	0	1	1 / 1 / 1
Plan, construct, design/draw	0.045	0.139	0	0.936	1 / 1 / 1
Instruct and hire employees	0.010	0.033	0	0.254	3 / 3 / 3
Dispose, organize, lead/lead	0.079	0.093	0	0.551	3 / 1 / 1
Educate/teach/train, advise	0.095	0.248	0	1	3 / 1 / 3
Jurisdiction, administration of justice	0.009	0.083	0	0.775	3 / 1 / 3
Care/supply, medical/cosmetic	0.073	0.198	0	0.900	3 / 1 / 3
Publish, work artistically	0.036	0.142	0	0.975	3 / 3 / 1
N(Occupations)	87				

Notes: [a]Occupational average of task item as listed in SJMM 1999 - 2016 at the 3-digit level of 87 Swiss occupations observed in the SJMM. [b]Task assignment of three independent experts: 1=complementary task, 2=substitutable task, 3=unaffected task.

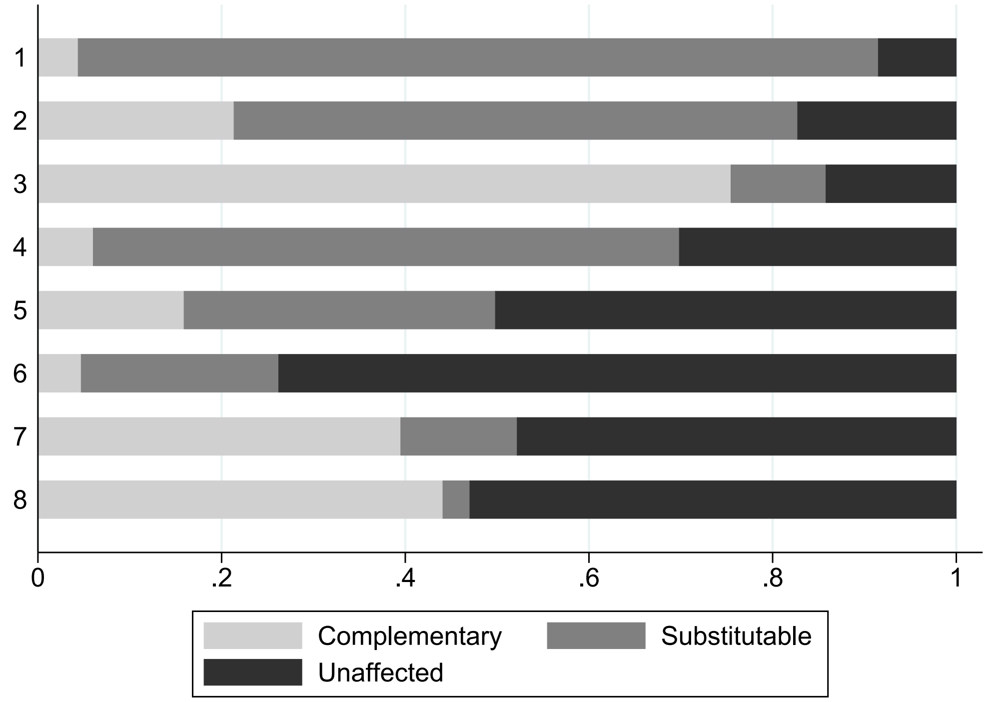
framework of Autor *et al.* (2003), and they assign the observed task items to three broad categories. The task assignment in the present chapter relies on the expertise of three labor market economists at the Swiss Federal Institute for Vocational Education and Training (SFIVET).

Specifically, I proceeded as follows: (i) every expert independently assessed each of the 21 SJMM task items to one of the three task types introduced in Section 3.2: complementary, substitutable, and unaffected. (ii) After taking the mean of these experts' assignments, every SJMM task item, and thus every job advertisement, is either considered as completely or partly complementary, substitutable, or unaffected *to* or *by* new technologies, respectively (Table 3.3).[8] (iii) I aggregate the three task types among the 43,932 job ads in my sample at the 3-digit SSCO 2000 level. Hence, every one of the 87 occupations observed in the SJMM consists of a complementary, a substitutable, and an unaffected task share that sum up to one.

Figure 3.3 illustrates that substitutable tasks represent a majority in agricultural (SSCO-1-digit: 1) and industrial occupations (2 and 4), while complementary tasks are frequent in IT and technical occupations (3). Unaffected tasks dominate in the remaining occupations, together with sizeable shares of substitutable tasks in trade (5) and hospitality occupations (6) and sizeable shares of complementary tasks in consulting (7) and social occupations (8). Aggregating over the whole sample, Table 3.1 shows that substitutable tasks account for 42% of all tasks among learned occupations, whereas complementary and unaffected tasks are less relevant across learned occupations, with shares of 37% and 21%, respectively. Current occupations, however, display higher shares of complementary tasks (44%) and unaffected tasks (26%) at the expense of substitutable tasks (30%).

[8] In fact, 9 task items were assigned to the same category by all three experts, while the other 12 items were assigned to two different categories, and none of the items was assigned to three distinct categories.

Figure 3.3: Task shares per 1-digit occupation



Notes: The figure shows the task shares at the level of eight SSCO 1-digit occupations. The calculation of the task shares is explained in Section 3.3.3. Occupations are: 1 Agricultural and forestry professions, livestock breeding professions, 2 Production occupations in industry and trade (excluding construction), 3 Technical and IT professions, 4 Professions in the building and construction industry and mining, 5 Trade and traffic professions, 6 Occupations in the hospitality industry and professions for the provision of personal services, 7 Professions in management and administration, banking, insurance and law, 8 Health, teaching and cultural professions, scientists.

3.4 Empirical strategy

3.4.1 Formalization of the hypothesis

To estimate the effect of horizontal mismatch on wages, I start with the following equation:[9]

$$log(Wage_{it}) = \alpha + \beta D_{it} + \gamma x_{it} + \psi z_{f[it]} + \phi_i + \theta_t + \epsilon_{it}$$

[9]For the sake of simplicity, I will refer to mismatch as a binary measurement in this Section. However, most estimates presented in Section 3.5 use the occupational distance as the main independent variable; Section 3.3.2 comprises the argumentation behind this.

$Wage_{it}$ measures individual i's net monthly wage at time t. D_{it} is the mismatch dummy, which takes a value of 1 if person i's current occupation is not equal to any of this person's learned occupation(s) on the 3-digit SSCO 2000 level at time t. Individual i's characteristics entering x_{it} are age, age-square, a dummy for being foreign, a dummy for having children, a dummy for being married, a categorical variable for the main three linguistic regions of Switzerland (German, French, Italian), a dummy for receiving further education in the past year, firm tenure and its square term, degree of employment in percent, dummies for being in a director or supervisor position, and a dummy for having a temporary contract. The term $z_{f[it]}$ captures the size of the firm f that individual i is employed at at time t by seven categories and an industry dummy for firm f (total twelve industries). The term ϕ_i represents unobserved person-fixed effects for person i and the term θ_t time-fixed effects for year t. Due to human capital losses in case of a horizontal mismatch, I expect a negative association between individuals' mismatch dummy D_{it} and their $wage_{it}$, i.e. $\beta < 0$.

3.4.2 Potential sources of bias

OLS estimations derived from the equation shown above may suffer from three sources of bias: heterogeneous horizontal mismatches, heterogeneity in unobserved person fixed effects, e.g. ability, and optimizing behavior in switching jobs.

First, the degree of human capital loss between someone's learned and this person's current occupation may vary for two reasons: (i) The loss of human capital generally increases with the occupation specificity of the mismatched person's human capital (Nordin *et al.*, 2010; Zhu, 2014). For instance, a person with a degree in medicine, which is considered to provide very specific occupational skills, can on average transfer less human capital when switching jobs than a person with a degree in business administration. (ii) The loss of human capital increases with the occupational distance between the learned and the current occupation of the mismatched person. For instance, an engineer who works as a technician loses much less of his human capital than if he worked as an office clerk. It therefore seems likely that people with more general skills are overrepresented in the mismatched population and that they choose occupations whose skill requirements are close to those of their learned occupations. The continuous mismatch measurement introduced in Section 3.3.2 accounts for this potential source of bias.

Second, unobserved personal characteristics ϕ_i likely affect mobility in the labor market. The most prominent example in the literature is ability. Gibbons and Katz (1991) show how individuals adversely select into job changes when outside employers cannot observe their ability. The same mechanism might lead to adverse selection of less able workers into mismatch (Boudarbat and Chernoff, 2012; Kucel and Vilalta-Bufi, 2012). However, this could also be reversed: high-ability workers might receive outside offers from firms regardless of their formal qualifications. Either way, an endogeneity bias potentially arises from individuals' unobserved ability. Formally, this biases β in the equation above upward if $cov(D_{it}, \phi_i > 0)$ and downward if $cov(D_{it}, \phi_i < 0)$.[10] Note that an upward bias corresponds to underestimating the negative wage effect and vice versa, since I presume $\beta < 0$. The mismatch literature usually deals with this second source of bias by applying fixed-effect estimations.

Third, wage offers arguably strongly affect individuals' job search behavior and labor market mobility (Mortensen, 1986; Rogerson *et al.*, 2005). Thus, simultaneous changes in the dependent variable, wage, and the independent variable of interest, the mismatch indicator, give rise to endogeneity concerns. For example, it seems likely that a proportion of the mismatched individuals become mismatched exactly because they can realize a wage gain through becoming so. As a consequence, the concept of mismatch, with its negative connotation, may be misleading. Presumably, people do not hesitate to accept a job offer at a higher wage, despite the fact that the job offered does not match their formal qualifications. Supporting this claim, Robst (2007a) finds that employees selecting into mismatch for payment and promotion opportunities earn substantially more than

[10]Note: this holds only for $\beta < 0$, e.g. a mismatch wage penalty, and $\phi_i > 0$, e.g. ability.

their matched counterparts. This leads to an underestimation of mismatch wage penalties in the model shown above compared to a scenario in which individuals are randomly assigned to mismatches. Section 3.4.3 elaborates more on this issue and Section 3.4.4 proposes an IV strategy to deal with it.

3.4.3 Fixed-effect estimations

Given the panel structure of the SHP, fixed-effect estimations are one obvious and promising strategy for addressing the second source of bias described above. Fixed-effect estimations allow it to keep unobservable personal characteristics (e.g. ability) fixed when analyzing wage effects of mismatched individuals. Exploiting the same dataset as in the present chapter and applying fixed-effect estimations, Schweri *et al.* (2019) neither find wage penalties for males considering themselves as mismatched nor males objectively being mismatched. Bender and Heywood (2009) apply fixed-effect estimations to a panel dataset of US doctoral students and reveal a small wage penalty for males but no wage penalty for females.

However, I argue that wage effects thus identified are most likely average effects, which aggregate different patterns leading to mismatch or match situations together with their distinctive wage implications. I claim that this appears because fixed-effect estimations fail to account for the likely dependency between individuals' switches into or out of mismatches and their wage expectations. This endogeneity concern corresponds to the third source of bias described above. To understand this concern in the context of horizontal mismatches, I now describe the four possible interactions between switches into or out of a mismatch situation and individuals' wages (see also Table 3.4). Moreover, I set out the implications of these four interaction cases for an identification within a fixed-effect setting:

- A person works in her or his learned occupation. Then the person accepts a position in an occupation different from any of her or his learned occupations offering a higher wage. Thus, I refer to this case as becoming voluntarily mismatched.[11] Because I observe a wage increase

[11] Note: I define the terms "voluntary" and "involuntary" in a purely monetary way. A

Table 3.4: Four mismatch cases

Share of task j in learned occ. $T_{it}^{j,locc}$	Case 1 *Voluntary mismatch*	Case 2 *Involuntary mismatch*	Case 3 *Voluntary match*	Case 4 *Involuntary match*
$j = complementary$	0.386	0.322	0.408	0.360
	(0.302)	(0.290)	(0.288)	(0.274)
$j = substitutable$	0.413	0.486	0.375	0.447
	(0.336)	(0.336)	(0.314)	(0.320)
$j = unaffected$	0.201	0.192	0.217	0.193
	(0.237)	(0.220)	(0.222)	(0.214)
$OccDist_i$ in t	0.989	1.044	–	–
	(0.414)	(0.526)		
$OccDist_i$ in $t-1$	–	–	0.763	0.771
			(0.441)	(0.391)
Number of cases	122	76	83	52
Share of all cases	0.366	0.228	0.249	0.156
$D_{i,t} - D_{i,t-1}$	$+1$	$+1$	-1	-1
$Wage_{i,t} - Wage_{i,t-1}$	> 0	< 0	> 0	< 0
$E[Wage_{i,t} \mid D_{i,t} = 1]$ $-E[Wage_{i,t-1} \mid D_{i,t-1} = 0]$	$+$	$-$	$-$	$+$

Notes: The four cases are described in Section 3.4.
Sources: SHP 1999 - 2016, SJMM 1999 - 2016.

and the mismatch dummy D_i switches from 0 to 1, the estimated wage effect of becoming mismatched yielded by fixed-effect estimations is positive. In my sample, I observe 122 such cases, accounting for 36.6% of all switches from match to mismatch or vice versa.[12]

switch accompanied by a wage increase is considered voluntary and a switch accompanied by a wage decrease is considered involuntary. Obviously, this omits other potential reasons for an occupational switch, such as any intrinsic motivation.

[12]It is possible that in some cases individuals underwent unemployment spells while switching from match to mismatch or vice versa. In this case, the switch from the occupation before to the occupation after the unemployment spell is relevant. The

- A person works in her or his learned occupation. Then the person loses her or his position and accepts a position in an occupation different from any of her or his learned occupations and offering a lower wage. Thus, I refer to this case as becoming involuntarily mismatched. Because I observe a wage decrease and the mismatch dummy D_i switches from 0 to 1, the estimated wage effect of becoming mismatched yielded by fixed-effect estimations is negative. I observe 76 such cases in my sample, accounting for 22.8% of all switches from match to mismatch or vice versa.

- A person works in an occupation different from any of her or his learned occupations. Then the person accepts a position in any of her or his learned occupations offering a higher wage. Thus, I refer to this case as becoming voluntarily matched. Because I observe a wage increase and the mismatch dummy D_i switches from 1 to 0, the estimated wage effect of becoming mismatched yielded by fixed-effect estimations is negative. I observe 83 such cases in my sample, accounting for 24.9% of all switches from match to mismatch or vice versa.

- A person works in an occupation different from any of her or his learned occupations. Then the person loses her or his position and accepts a position in any of her or his learned occupations offering a lower wage. Thus, I refer to this case as becoming involuntarily matched. Because I observe a wage decrease and the mismatch dummy D_i switches from 1 to 0, the estimated wage effect of becoming mismatched yielded by fixed-effect estimations is positive. I observe 56 such cases in my sample, accounting for 15.6% of all switches from match to mismatch or vice versa.

In a fixed-effect setting, Cases 2 and 3 reveal a mismatch wage penalty, while Cases 1 and 4 reveal a wage increase for being mismatched. Con-

unemployment spell is simply omitted because my sample only consists of employed individuals. However, Table 3.10 indicates that the coefficients yielded by my wage estimations are insensitive to the inclusion of individuals undergoing an unemployment spell during the sample period.

sequently, any overall effect provided by fixed-effect estimations merely represents an effect on average that is too small for one subpopulation and wrongly identified for another subpopulation. Conclusions derived from these fixed-effect estimations are therefore potentially misleading. If, for example, the effect of a mismatch wage penalty is mainly driven by graduates who find their first position in their learned occupation (Case 3), one cannot derive any conclusion, or one may even derive a false conclusion, for individuals who are involuntarily mismatched due to a layoff (Case 2). Moreover, fixed-effect estimations are based on individuals switching from match to mismatch or vice versa during the sample period. In total, I merely observe 333 (198 to mismatch and 135 to match) such switches stemming from 248 individuals. In contrast, individuals being mismatched (n=525) or matched (n=457) throughout the entire sample period do not contribute to any effect yielded by fixed-effect estimations. This is especially worrying if the subpopulation of individuals mismatched or matched throughout the entire sample period differs systematically from individuals switching within the sample period.

3.4.4 IV approach

This section introduces an IV approach to address the sources of bias described in Section 3.4.2 and the remaining shortcomings of fixed-effect estimations described in Section 3.4.3. The first part of the section discusses the choice of the instrument and its hypothesized first-stage relation to the endogenous mismatch variable. In addition to this first-stage relation, an instrument needs to satisfy the independence assumption and the exclusion restriction to be valid. These two requirements are not formally testable and thus demand a critical examination (Angrist and Pischke, 2014; Imbens, 2014). The remaining parts of this section provide this.

First-stage

Leaning on the task-based approach, Section 3.2 highlights how technological change affects demand differently for three task types (complementary, substitutable, and unaffected) in the labor market, and therefore also for occupations bundling these tasks (Figure 3.1). The same Section 3.2 pointed out how task-shifting technological change potentially increases or decreases horizontal mismatches between demanded tasks and supplied skills in the labor market. Based on this general pattern, I propose two specific mechanisms that presumably underpin an association between the task shares of individuals' learned occupations and their mismatch probability.

First, *interoccupation* employment shifts lower the number of positions in occupations bundling substitutable tasks, while positions in occupations bundling complementary tasks become widely available. These interoccupation shifts result in an oversupply of individuals with learned substitutable occupations and thus trigger mismatches among them. However, numbers of positions in occupations bundling complementary or unaffected tasks increase or remain stable. This leads to low mismatch numbers among individuals who learned these occupations.

Second, demand rises for complementary tasks within occupations (*intraoccupation*) at the expense of substitutable tasks. Consequently, returns to complementary tasks rise, while returns to substitutable tasks decline across occupations.[13] This enhances varying occupational mobility patterns depending on the task composition of individuals' learned occupations. On the one hand, the increasing demand for mostly complementary skills across various occupations provides well-paid open positions for individuals with these complementary although they lack the formal qualifications for these occupations. This might pull individuals into mismatches which are, however, perceived as voluntary.[14] On the other hand, the demand for skills executing substitutable tasks deceases even within occupations bundling these tasks to a large extent. This lowers the demand for individuals with mostly substitutable skills both generally and within their learned occupations. Thus, intraoccupational task shifts might push these individuals out of their

[13]Spitz-Oener (2006) and Dengler and Matthes (2015) described these intra-occupation task shifts among German occupations.

[14]Early works underpinning how occupational mobility does not necessarily lead to wage decreases, even though movers are expected to lose some of their specific human capital, include Johnson (1978), Topel and Ward (1992), and Neal (1999).

learned occupations and lead to mismatches that are perceived as involuntary. Neither a pull- nor a push-mechanism concerns individuals in primarily unaffected learned occupations. Demand for their skills outside their learned occupations and thus the number of outside options pulling them out of their learned occupation remains stable. Meanwhile, the task composition of their learned occupations does not shift towards more complementary tasks. This limits the attractiveness of their learned occupations for individuals with a more complementary skill bundle and therefore also their potential for being pushed out of their learned occupation.

Considering the continuous mismatch measurement introduced in Section 3.3.2, both of these effects on the extensive margin, being mismatched, reinforce themselves on the intensive margin, which accounts for the magnitude of a perceived mismatch in terms of occupational distance. Individuals who learned complementary occupations are, if mismatched, likely to find a position in a close occupation, due to a similar occupational task composition, for which demand also increased. In contrast, lower demand for occupations bundling substitutable tasks might force individuals who learned these occupations to move to rather different occupations if they are mismatched.

Overall, I primarily expect task-shifting technological change to increase mismatches among individuals who learned occupations bundling substitutable tasks. Second, task-shifting technological change triggers two opposing effects for individuals who learned rather complementary occupations, and it remains a priori ambiguous which effect dominates. Third and finally, as by definition technology has little impact on unaffected tasks, I assume individuals who learned occupations bundling these tasks are seldom prone to mismatches. These hypothesized associations between the task composition of individuals' learned occupations and their mismatch incidence represent the first-stage of my IV setting. Section 3.5.1 applies OLS estimations to evaluate them.

Independence assumption

The independence assumption requires the instrument to be randomly or as good as randomly assigned (Angrist and Pischke, 2008). The first-best solution to meet this requirement is by design. Prominent examples of such instruments are draft lotteries (e.g. Angrist, 1990), giving birth to twins (Angrist *et al.*, 2010; Rosenzweig and Wolpin, 1980), and random variation in newborns' gender composition (Angrist and Evans, 1998). In contrast, the task share of someone's learned occupation is obviously not randomly assigned. For example, substitutable tasks often bundle in blue-collar occupations, which are accessible without a university degree. Hence, the task share of someone's learned occupation correlates with education and affects the dependent variable wage.

Thus, I need to relax the random assignment assumption and require it "to hold only within subpopulations defined by covariates" (Imbens, 2014, p.27).[15] Optimally, these covariates include all factors that affected an individual's occupational choice and are thus determined prior to the instrument (Deuchert and Huber, 2017). One such factor is a person's gender, assuming males choose more manual occupations ceteris paribus. However, this control is redundant due to the sample's restriction to males only. Another factor determining someone's occupational choice is school performance. Most occupations bundle people of similar school performances, and many occupations require a certain level of education. Thus, I would like to control for a person's school performance prior to any occupational choice, for instance at the end of compulsory school. Unfortunately, the Swiss Household Panel does not contain this information. The strategy to mitigate the issue is twofold.

In all estimations, I control for different educational categories. These educational attainment categories aim to approximate an individual's school performance prior to his or her occupational choice. I surmise that the task share of an individual's chosen occupation is as good as random once I control

[15] Baiocchi *et al.* (2010) use matching methods for the same purpose.

for educational attainment.[16]

In Section 3.5.3, I present IV estimations within educational subgroups. The motivation for this is similar to that above: the task share of someone's learned occupation is plausibly closer to random within youngsters opting for VET (e.g. becoming a commercial clerk or an electrician) than between a youngster opting for VET and a youngster applying for university. Moreover, these subsample estimations permit to identify varying mismatch wage penalties across different educational cohorts.

Exclusion restriction

The exclusion restriction requires that the instrument affects the outcome only through the endogenous variable (Angrist and Pischke, 2008). In a violation of this restriction, it seems likely that the task share of individuals' learned occupations directly affects their wages. Specifically, I assume the channel through which the task content of individuals' learned occupations affects their current wages works through the task content of their current occupations.[17] This enables this direct link to be clipped by controlling for the share of task j of individuals' current occupations *cocc* at time t, $T_{it}^{j,cocc}$; whereby the share of task type j of individuals' learned occupations *locc* at time t, $T_{it}^{j,locc}$ represents the instrument. Deuchert and Huber (2017) underpin theoretically how the exclusion restriction can be sustained by controlling for any direct effects.

Taking these requirements into account and replacing the mismatch dummy D_{it} with the preferred continuous mismatch measurement $OccDist_{it}$,

[16]Obviously, this is a rather strong claim and I thus encourage attempts to measure how individuals are affected by task-shifting technological change in a way that is more endogenous to their occupational choice.

[17]Note that I already set out above why controlling for educational attainment, among other factors, is necessary in the present IV setting. Though I argued this is the case due to the nonrandom assignment of the instrument, one could also argue that education is another channel through which individuals' learned occupations affect their wages directly. In this sense, controlling for educational attainment also helps clip any direct link between the instrument and the outcome variable and thus satisfy the exclusion restriction.

the equation shown above transforms into the two following equations:

$$log(Wage_{it}^{cocc}) = \beta_0 + \beta_1 \widehat{OccDist}_{it} + \beta_2 x_{it} + \beta_3 z_{f[it]} + \beta_4 T_{it}^{j,cocc} + \epsilon_{it},$$

where $\widehat{OccDist}_i$ is instrumented as follows:

$$OccDist_{it} = \alpha_0 + \alpha_1 T_{it}^{j,locc} + \alpha_2 x_{it} + \alpha_3 z_{f[it]} + \alpha_4 T_{it}^{j,cocc} + \upsilon_{it}$$

$Wage_{it,cocc}$ is individual i's wage in current occupation *cocc* at time t. As set out above, individual characteristics x_{it} crucially include educational dummies (VET, VET high school, VET high school with baccalaureate, technical or vocational school, Tertiary-B track, university of teacher education, universities of applied science, universities, and post-graduate degrees). Adding the share $T_{it}^{j,cocc}$ of task j of individual i's current occupation *cocc* at time t sustains the exclusion restriction. The occupational distance $OccDist_{it}$ is now instrumented with the respective share of task j of an individual i's learned occupation *locc* at time t, $T_{it}^{j,locc}$.[18]

3.5 Results

3.5.1 First-stage results

Table 3.5 exploits the association between the task shares of individuals' learned occupations and their mismatch incidence in three ways. First, columns (1) to (3) regress the mismatch dummy D_{it} on the task shares of individuals' learned occupations (extensive margin). Second, columns (4) to (6) restrict the sample to mismatched individuals and yield the correlation between the task composition of their learned occupations and the strength of their mismatch in terms of occupational distance $OccDist_{it}$ (intensive margin). Third, columns (7) to (9) display the first-stage estimates by regressing the preferred mismatch measurement $OccDist_{it}$ on the task shares of individuals' learned occupations. I now discuss these first-stage relations

[18] Note that individuals' learned occupations are also potentially time-variant due to new formal qualifications acquired during the sample period.

together with the three hypotheses concerning their direction presented in Section 3.4.4.

According to column (1), the association between the complementary task share of individuals' learned occupations and their mismatch probability is rather weak and statistically only significant at the 10%-level. This seems to support the ambiguous relation between the complementary task share of individuals' learned occupations and their mismatch probability hypothesized in Section 3.4.4. Individuals with complementary learned occupations are able to avoid mismatches due to broadly available positions within their learned occupation. However, these individuals also profit from an increased demand for their skills in other occupations, which might allow them to realize a wage gain while becoming mismatched. In contrast, and in line with the hypothesis presented in Section 3.4.4, the substitutable task share of individuals' learned occupations and their mismatch probability correlates positively (column 2). According to the point estimate in column (2), a one-standard-deviation (0.26) higher substitutable task share is associated with a higher mismatch probability of 15 percentage points. Presumably, the scarcity of positions in occupations bundling substitutable tasks augments the mismatch probability of individuals who learned these occupations. The negative coefficient in column (3) suggests the opposite for individuals who learned rather unaffected occupations.

In columns (4) to (6), I regress individuals' occupational distance $OccDist_{it}$ on the three types of task shares of someone's learned occupations while restricting the sample to mismatched individuals, $OccDist_{it} > 0$. This allows me to interpret the resulting estimate as the first-stage association on the intensive margin. According to column (5), individuals with substitutable learned occupations struggle to find close occupations in case of a mismatch. This seems plausible, because these close occupations likely bundle mostly substitutable tasks as well and therefore face the same decline in demand. In contrast, columns (4) and (6) suggest that mismatched individuals with complementary and to an ever greater extend unaffected learned occupations tend to move to close occupations in terms of $OccDist_{it}$. This seems consistent with the hypothesis that positions in occupations

Table 3.5: First-stage estimates

	D_{it}			$OccDist_{it}$ if $D_{it} = 1$			$OccDist_{it}$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$T_{it}^{j,locc}$, $j = comp.$	-0.121*			-0.225***			-0.225**		
	(0.073)			(0.068)			(0.100)		
$T_{it}^{j,locc}$, $j = subs.$		0.553***			0.661***			1.113***	
		(0.074)			(0.103)			(0.126)	
$T_{it}^{j,locc}$, $j = unaff.$			-0.574***			-0.572***			-1.203***
			(0.098)			(0.127)			(0.161)
Constant	0.112	0.239	0.098	0.581**	0.263	0.339	-0.032	-0.003	-0.146
	(0.212)	(0.206)	(0.209)	(0.290)	(0.305)	(0.267)	(0.266)	(0.244)	(0.245)
Personal controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wave Dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$T_{it}^{j,cocc}$	$j = comp.$	$j = subs.$	$j = unaff.$	$j = comp.$	$j = subs.$	$j = unaff.$	$j = comp.$	$j = subs.$	$j = unaff.$
R-squared	0.083	0.159	0.113	0.165	0.229	0.282	0.080	0.238	0.204
Observations	10471	10471	10471	5483	5483	5483	10471	10471	10471
F-stat	2.73	55.94	34.56	10.98	41.23	20.18	5.10	78.49	55.99

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses clustered at individual level. $T_{it}^{j,cocc}$ refers to the task share of an individual's current occupation of the same task category as the one that is the independent variable in the respective model.

bundling complementary or unaffected tasks become widely available or at least remain stable. However, the lower coefficient in column (4) than in column (6) somewhat contradicts this argument. One explanation could be that individuals with complementary learned occupations disproportionately possess skills that are demanded throughout various occupational fields. Therefore, these individuals receive more attractive job offers in occupations that are relatively unrelated to their learned occupations than do individuals with unaffected learned occupations.

Accounting for the association between the task share of individuals' learned occupations and their mismatch probability on both the extensive margin and the intensive margin simultaneously yields the preferred first-stage estimates in columns (7) to (9). Overall, the displayed coefficient in column (8) indicates high occupational distances for individuals with learned occupations bundling substitutable tasks, while individuals with learned occupations bundling tasks unaffected by new technology display smaller occupational distances in column (9). To be precise, an additional standard deviation in the share of substitutable tasks in individuals' learned occupation increases their occupational distance by 29.3 percentage points, whereas the equivalent for the share of unaffected tasks decreases occupational distance by 23.4 percentage points. In contrast, the correlation between the complementary task share of individuals' learned occupation and their occupational distance is about five times smaller and statistically less significant (column 7).

Considering the general rule that F-statistics above 20 indicate sufficient explanatory power at the first-stage (Bound *et al.*, 1995), only the share of substitutable and the share of unaffected tasks are valid instruments, while the share of complementary tasks has insufficient explanatory power for individuals' mismatch probability. In forthcoming analysis, I therefore employ the share of substitutable and unaffected tasks as instruments and include them simultaneously in most estimations.

3.5.2 Main results

Occupational distance

Table 3.6 presents the main results relying on the continuous mismatch measurement (see Section 3.3.2) as the dependent variable.[19] The table displays OLS, fixed-effect, and IV (first-stage and 2SLS) estimations as described in Section 3.4.

The OLS estimate in column (1) of Table 3.6 reveals a very small positive average wage difference of 0.7% between a matched and an average mismatched person in terms of occupational distance, all else being equal. The fixed-effect estimate in column (2) accounts for unobserved heterogeneity among individuals selecting into mismatch and is slightly negative but close to zero as well. This suggests that no wage penalty stems from becoming horizontally mismatched. However, as highlighted in Section 3.4, I argue this fixed-effect estimate merely yields an average effect by mixing different underlying mechanisms pointing in distinct directions.

Thus, I draw attention to the IV estimation results in columns (3) to (8) of Table 3.6. Column (4) utilizes the share of substitutable tasks as an instrument and yields a mismatch wage penalty of 11.9% for a person in an average mismatch in terms of occupational distance.[20] This estimate is statistically significant at the one percent level and economically relevant. With an annual gross wage of roughly 117,000 (median: 105,000) Swiss francs, this amounts to 13,900 (12,500) Swiss francs wage penalty per year.

In contrast, column (6) uses the share of unaffected tasks as an instrument to yield a smaller mismatch wage penalty (-6.2%) that is statistically insignificant. This result arises due to the relatively small and imprecise reduced-form estimate exploiting the unaffected task share displayed in

[19]Table A.1 contains an extended output and shows the coefficients of various control variables.

[20]Note that the median of the occupational distance measurement for the mismatched individuals is 0.96 and thus very close to the mean of one. In conclusion, the effective mismatch penalty exceeds the estimated 11.9% for roughly half of all mismatched individuals and falls behind this 11.9% for the other half of all mismatched individuals.

Table 3.6: Main results – continuous mismatch

	$Wage_i$							
			IV		IV		IV	
	OLS	FE	First-stage	2SLS	First-stage	2SLS	First-stage	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$OccDist_{it}$	0.007	-0.007		-0.119***		-0.062		-0.098***
	(0.011)	(0.012)		(0.033)		(0.043)		(0.029)
$T_{it}^{j,locc}$ $j = subs.$			1.113***				0.957***	
			(0.126)				(0.107)	
$T_{it}^{j,locc}$ $j = unaff.$					-1.203***		-0.874***	
					(0.161)		(0.149)	
Constant	8.983***	8.643***	-0.003	9.151***	-0.146	8.912***	-0.141	9.108***
	(0.104)	(0.116)	(0.244)	(0.108)	(0.245)	(0.106)	(0.234)	(0.105)
Personal controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Educational dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$T_{it}^{j,cocc}$	No	No	$j = subs.$	$j = subs.$	$j = unaff.$	$j = unaff.$	$j^1 = subs.$	$j^1 = subs.$
							$j^2 = unaff.$	$j^2 = unaff.$
Instrument, $T_{it}^{j,locc}$	–	–	–	$j = subs.$	–	$j = unaff.$	–	$j^1 = subs.$
								$j^2 = unaff.$
R-squared	0.562	0.435	0.238	0.562	0.204	0.554	0.314	0.571
Observations	10471	10471	10471	10471	10471	10471	10471	10471
F-stat(first-stage)			78.49		55.99		52.24	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses clustered at individual level. $T_{it}^{j,cocc}$ refers to the task share of an individuals current occupation of the same task category as the one that is used as an instrument in the respective model.

column (3) of Table A.2.

How can these different mismatch wage penalties shown in Table 3.6 be explained?[21] Following Imbens (2014) and Angrist and Pischke (2014), I interpret the negative effect yielded in column (4) and the small and imprecisely estimated effect yielded in column (6) as representing local average treatment effects (LATE) for different subgroups, or *compliers* as Imbens (2014) and Angrist and Pischke (2014) call them. Compliers suffering from the mismatch wage penalty yielded in column (4) are mismatched individuals with highly substitutable learned occupations who would not be mismatched had they learned less substitutable occupations. Task shifting technological change increased their mismatch probability and their occupational distance in case of a mismatch by shortening available positions both within their learned occupations and within close occupations. Moreover, a general shift away from substitutable tasks limits their returns to skills needed to execute these tasks across occupations. Therefore, individuals with mostly substitutable learned occupations (compliers in column 4 of Table 3.6) receive relatively poor outside options and thus suffer from a mismatch wage penalty.

Conversely, the LATE yielded in column (6) of Table 3.6 pertains to individuals with learned occupations bundling few unaffected tasks.[22] However, it remains ambiguous how returns evolved for these other-than-unaffected-tasks in the general labor market and thus how outside options for these compliers created. Hence, having few skills to execute unaffected tasks and being mismatched is not sufficient to suffer from a mismatch wage penalty. Moreover, these compliers arguably represent a mixture of individuals with complementary and substitutable learned occupations.

[21] From an econometric point of view, this is straightforward: The IV estimate equals the ratio of the reduced form to the first-stage estimate. While the first-stage estimates are both statistically significant different from zero (Table 3.5), the reduced form estimates of Table A.2 display a negative and statistically significant association between $T_{it}^{locc,j}$ and $wage_{it}$ for $j = substitutable$ but not for $j = unaffected$.

[22] They comply with being mismatched by having learned occupations bundling few unaffected tasks, whereas they would not be mismatched had they learned more unaffected tasks.

In Section 3.2, I argued that the former likely select into mismatches due to valuable outside options whereas the latter tend to be mismatched involuntarily. Arguably, this further contributes to the insignificant reduced-form association between the share of unaffected tasks of individuals' learned occupations and their wages (column 3 of Table A.2), and thus to the statistically insignificant mismatch wage effect for this subgroup in column (6) of Table 3.6. However, one should notice that the 95% confidence intervals of the point estimates in columns (4) and (6) overlap. Thus, one cannot reject the null hypothesis that the two estimated effects are statistically not different.

In column (8), I exploit the shares of substitutable and unaffected tasks simultaneously as instruments. The mismatch wage penalty revealed amounts to roughly 10%. Again, at the 95% level, this estimate is not significantly different from the point estimates in columns (4) and (6). Column (7) reveals the corresponding first-stage and points to the same conclusion as columns (3) and (5): having learned a complementary occupation is positively correlated with occupational distance, while having learned an unaffected occupation is associated with a low occupational distance between learned and current occupations.

Mismatch dummy

Table 3.7 employs the classical binary mismatch variable. Consistent with the higher average wages among mismatched individuals yielded in Table 3.1, the OLS estimate displays a positive correlation between the mismatch dummy and wages (column 1, Table 3.7). However, this positive association vanishes when applying fixed-effect estimations and thus accounting for unobservable personal characteristics in column (2) of Table 3.7.

Turning to the IV estimations in columns (3) to (8), the much higher coefficient using the binary mismatch variable in Table 3.7 compared to the same specifications in Table 3.6 is striking. It thus seems questionable how credible a mismatch wage penalty of 24% is (column 4, Table 3.7). The underlying reduced form estimations in model (4) of Table 3.7 do not differ

Table 3.7: Main results – mismatch dummy

	$Wage_i$							
			IV		IV		IV	
	OLS	FE	First-stage	2SLS	First-stage	2SLS	First-stage	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
D_{it}	0.033**	-0.005		-0.240***		-0.130		-0.205***
	(0.013)	(0.014)		(0.070)		(0.094)		(0.063)
$T_{it}^{j,locc}$ $j = subs.$			0.553***				0.477***	
			(0.074)				(0.073)	
$T_{it}^{j,locc}$ $j = unaff.$					-0.574***		-0.403***	
					(0.098)		(0.102)	
Constant	8.980***	8.645***	0.239	9.209***	0.098	8.934***	0.210	9.166***
	(0.104)	(0.116)	(0.206)	(0.120)	(0.209)	(0.109)	(0.206)	(0.115)
Personal controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Educational dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$T_{it}^{j,cocc}$	No	No	$j = subs.$	$j = subs.$	$j = unaff.$	$j = unaff.$	$j^1 = subs.$	$j^1 = subs.$
							$j^2 = unaff.$	$j^2 = unaff.$
Instrument, $T_{it}^{j,locc}$	–	–	–	$j = subs.$	–	$j = unaff.$	–	$j^1 = subs.$
								$j^2 = unaff.$
R-squared	0.563	0.435	0.159	0.496	0.113	0.526	0.178	0.520
Observations	10471	10471	10471	10471	10471	10471	10471	10471
F-stat(first-stage)			55.94		34.56		37.29	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses clustered at individual level. $T_{it}^{j,cocc}$ refers to the task share of an individuals current occupation of the same task category as the one that is used as an instrument in the respective model.

from model (4) of Table 3.6; in both cases the substitutable task share of an individual's learned occupation is regressed on his or her wage. Consequently, this difference is entirely driven by the underlying first-stage estimations.[23] Table 3.5 displays these diverging first-stage estimates for the binary (D_{it}, column 2, Table 3.5) and the continuous ($OccDist_{it}$, column 8, Table 3.5) mismatch measurement. Note the much lower R^2 in column (2) of Table 3.5 compared to column (8) of Table 3.5. Apparently, the substitutable task share of an individual's learned occupation contains more explanatory power for mismatches measured continuously than captured by a binary variable. Again, I conclude that the use of the continuous mismatch measurement adds additional important information that the the mismatch dummy misses.

3.5.3 Subsample results

Education cohorts

Recently, Hanushek *et al.* (2017) argued that vocational education's advantages in smoothing the school-work transition at the beginning of individuals' careers turns into a disadvantage in both employment and wages at later career stages. One mechanism postulated by Hanushek *et al.* (2017) to explain this pattern is the higher specificity of human capital acquired during a vocational education compared to the rather general human capital acquired at a high school. Individuals with a vocational degree easily find a position directly after graduation but face difficulties adjusting their specific human capital to changing demands in the labor market occurring during their career, e.g. due to technological change. Conversely, individuals with general human capital often face difficulties bringing their general skills to a specific position right after graduation, but they adapt more easily to a changing labor market and thus their employment and wage perspectives develop well over their lifecycle.[24] Horizontal skill mismatches

[23]Again, the IV estimate equals the ratio of the reduced form to the first-stage estimation.

[24]Partly supporting these hypotheses, Korber and Oesch (2019) find substantially lower earnings for Swiss individuals with a vocational degree once they enter their thirties.

among individuals with a vocational degree might be one channel through which Hanushek *et al.*'s (2017) arguments materialize.

Table 3.8 tests this claim by estimating mismatch wage penalties among education cohort (vocational education and training VET, Tertiary-B, and Tertiary-A) subsamples.[25] All OLS and fixed-effect estimations show close-to-zero and statistically insignificant effects as in the full sample estimations in Table 3.6. IV estimations in columns (3), (6), and (9) rely on the share of substitutable tasks as an instrument that yielded a statistically significant mismatch wage penalty in the full sample (column 4 of Table 3.6).

Overall, the findings presented in Table 3.8 suggest that horizontal mismatches triggered by task-shifting technological change do not represent a channel through which Hanushek *et al.*'s (2017) argument materializes. Although the mismatch wage penalty for individuals with a VET degree is somewhat smaller than for their counterparts with a tertiary degree, these differences are statistically not significant. Furthermore, one should note the relative weak first-stage association among individuals with a tertiary-A degree, which leads to the very imprecise IV estimate in column (9).

Age cohorts

Table 3.9 presents subsample estimations for individuals who are younger than 46 and individuals who are 46 or older. Simple conditional comparisons of matched and mismatched individuals in columns (1) and (4) reveal that mismatched individuals only earn less than their matched counterparts at later stages of their work careers. Interestingly, fixed-effect estimations controlling for heterogeneity in unobservable personal characteristics reveal

However, they do not find diverging employment chances across education cohorts over the lifecycle.

[25]VET refers to people who completed some sort of apprenticeship training. The Tertiary-B track represents a tertiary track that usually requires the prior completion of an apprenticeship. The Tertiary-A track includes universities, which are either accessible for high-school graduates or apprentices with a baccalaureate, and universities of applied science, which require an apprenticeship with a baccalaureate or a high school degree and some work experience in the field of study.

Table 3.8: Subsample results – education cohorts

	$Wage_i$								
	VET			Tertiary-B			Tertiary-A		
	OLS	FE	IV	OLS	FE	IV	OLS	FE	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$OccDist_{it}$	0.009	-0.007	-0.083*	-0.005	-0.000	-0.150***	-0.002	-0.013	-0.156
	(0.015)	(0.015)	(0.045)	(0.017)	(0.016)	(0.045)	(0.031)	(0.033)	(0.125)
Constant	8.967***	8.795***	9.199***	9.326***	8.762***	9.413***	8.996***	8.639***	9.015***
	(0.138)	(0.144)	(0.140)	(0.225)	(0.204)	(0.221)	(0.245)	(0.248)	(0.249)
Personal controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Educational dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$T_{it}^{j,cocc}$	No	No	$j = subs.$	No	No	$j = subs.$	No	No	$j = subs.$
Instrument, $T_{it}^{j,locc}$	–	–	$j = subs.$	–	–	$j = subs.$	–	–	$j = subs.$
R-squared	0.486	0.366	0.518	0.411	0.441	0.408	0.508	0.462	0.471
Observations	4012	4012	4012	3701	3701	3701	2758	2758	2758
F-stat(first-stage)			32.51			49.39			8.146

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses clustered at individual level. $T_{it}^{j,cocc}$ refers to the task share of an individuals current occupation of the same task category as the one that is used as an instrument in the respective model.

the opposite (columns 2 and 5). One explanation for this pattern might be that individuals' accumulated informal human capital dominates at later career stages over formally acquired degrees and thus leads to the diminishing of a mismatch wage penalty in the later stages of someone's career. Presumably, wage negotiations between firms and young job seekers are mostly based on formal degrees: What else can firms observe? As a consequence, young people suffer from a wage penalty if they cannot find a position matching their formal degree. At later stages of individuals' careers, firms and job seekers are increasingly able to base their wage offers on informally accumulated human capital, which becomes increasingly visible through job references and previous occupationally relevant performance. Therefore, firms are willing to pay relatively high wages irrespective of the formal qualification a job applicant has or has not.

The IV estimates are somewhat stronger among the older cohort (column 6 of Table 3.9) compared to the younger cohort (column 3 of Table 3.9). However, the confidence intervals of the estimates widely overlap. It seems that task-shifting technological change triggering mismatches is a threat for individuals at various stages of their career.

Unemployment spells

Task-shifting technological change arguably increases not only horizontal mismatches but also unemployment. However, observations during an unemployment spell drop out of my sample because I am unable to define the mismatch status of unemployed individuals. Assuming unemployment is financially more harmful than being mismatched, this leads to an underestimation of the overall financial losses suffered by individuals who lose a position in their previously learned occupation and become mismatched or unemployed. Table 3.10 tackles this concern by applying the main estimations presented in Table 3.6 to a subsample of individuals who were always employed when being observed. The results in Table 3.10 are almost identical to those in Table 3.6 and thus mitigate concerns that previously shown results are biased due to unemployment spells.

Table 3.9: Subsample results – age cohorts

	$Wage_i$					
	Age 25-45			Age 46-65		
	OLS	FE	IV	OLS	FE	IV
	(1)	(2)	(3)	(4)	(5)	(6)
$OccDist_{it}$	0.009	-0.030**	-0.111***	-0.043***	0.016	-0.140***
	(0.013)	(0.013)	(0.034)	(0.014)	(0.024)	(0.050)
Constant	8.541***	8.258***	8.608***	8.622***	9.089***	8.817***
	(0.153)	(0.178)	(0.157)	(0.580)	(0.435)	(0.623)
Personal controls	Yes	Yes	Yes	Yes	Yes	Yes
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes
Educational dummies	Yes	Yes	Yes	Yes	Yes	Yes
$T_{it}^{j,cocc}$	No	No	$j = subs.$	No	No	$j = subs.$
Instrument, $T_{it}^{j,locc}$	–	–	$j = subs.$	–	–	$j = subs.$
R-squared	0.596	0.545	0.587	0.514	0.145	0.481
Observations	5203	5203	5203	5268	5268	5268
F-stat			77.26			35.21

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses clustered at individual level. $T_{it}^{j,cocc}$ refers to the task share of an individuals current occupation of the same task category as the one that is used as an instrument in the respective model.

Table 3.10: Subsample results – exclude individuals with unemployment spell(s)

	$Wage_i$							
			IV		IV		IV	
	OLS	FE	First-stage	2SLS	First-stage	2SLS	First-stage	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$OccDist_{it}$	0.012	0.004		-0.119***		-0.056		-0.097***
	(0.011)	(0.012)		(0.033)		(0.043)		(0.029)
$T_{it}^{j,locc}$ $j = subs.$			1.113***				0.957***	
			(0.126)				(0.107)	
$T_{it}^{j,locc}$ $j = unaff.$					-1.203***		-0.874***	
					(0.161)		(0.149)	
Constant	9.011***	8.651***	-0.003	9.180***	-0.146	8.941***	-0.141	9.134***
	(0.107)	(0.120)	(0.244)	(0.109)	(0.245)	(0.108)	(0.234)	(0.106)
Personal controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Educational dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$T_{it}^{j,cocc}$	No	No	$j = subs.$	$j = subs.$	$j = unaff.$	$j = unaff.$	$j^1 = subs.$	$j^1 = subs.$
							$j^2 = unaff.$	$j^2 = unaff.$
Instrument, $T_{it}^{j,locc}$	–	–	–	$j = subs.$	–	$j = unaff.$	–	$j^1 = subs.$
								$j^2 = unaff.$
R-squared	0.564	0.452	0.238	0.564	0.204	0.557	0.314	0.574
Observations	9923	9923	10471	9923	10471	9923	10471	9923
F-stat(first-stage)			75.43		53.61		49.75	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses clustered at individual level. People with at least one unemployment spell are excluded. $T_{it}^{j,cocc}$ refers to the task share of an individuals current occupation of the same task category as the one that is used as an instrument in the respective model.

3.6 Conclusion

This chapter investigates whether horizontally mismatched individuals suffer from a wage penalty. The answer, on average and as revealed by OLS estimations, is no. Individuals who work in an occupation for which they have no formal degree earn, conditionally on observables, even more than similar individuals. When fixed-effect estimations are used, thus accounting for unobservable covariates such as ability, this positive association vanishes.

However, in the present chapter, I argue that this is not the end of the story. Descriptive evidence on individuals' wage evolution when becoming matched or mismatched suggests that fixed-effect estimations aggregate heterogeneous occupational mobility patterns and thus merely yield a mismatch wage effect on average. Presumably, some individuals select into mismatches to realize a higher wage, while others are negatively affected by a mismatch incidence and suffer from a wage penalty.

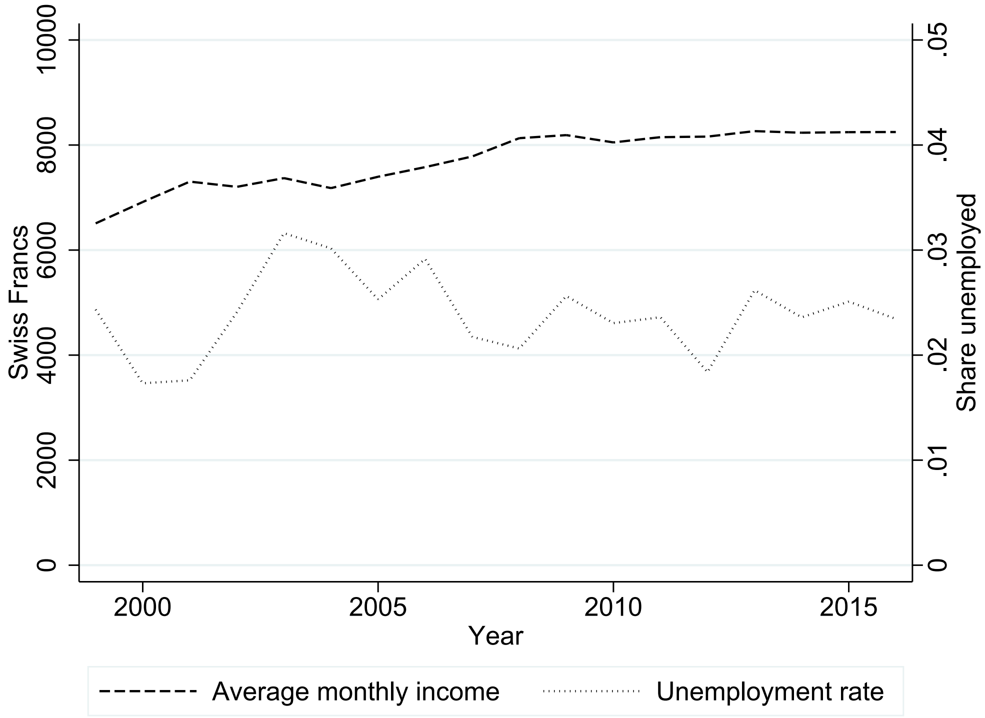
To cope with this heterogeneity and to isolate the most relevant mismatch incidences from a policy perspective, I propose an IV approach. In times of task-shifting technological change, I surmise that individuals' mismatch probability is positively correlated with the substitutability of their learned occupations' task bundle. In contrast, individuals who learned occupations bundling few affected tasks are more likely to stay in their learned occupation. Based on this pattern, I regard the task composition of individuals' learned occupations, conditional on their educational attainment and the task composition of their current occupation, as a valid instrument for the endogenous mismatch variable.

Applying this IV approach to a sample of roughly 10,500 person-year observations, I estimate a negative wage effect of roughly 12% for mismatched Swiss males. However, this a mismatch wage penalty is only revealed when exploiting the share of substitutable tasks across individuals' learned occupations as an instrument. This seems plausible: the estimated 12% mismatch wage penalty refers to a LATE for the subgroup of compliers who in this setting are mismatched individuals with learned occupations displaying high shares of substitutable tasks who would not be mismatched had they learned less substitutable occupations.

In conclusion, mismatches represent a labor market phenomenon with diverse aspects. Many individuals accept mismatches to increase their salaries. Based on classical economic theory, in which a wage increase amplifies a better employee-employer match in terms of human capital allocation, the term *mis*match is therefore often misleading. However, the analysis in this chapter shows that some mismatches are associated with wage losses. From a policy perspective, these mismatches might be the most relevant because they are not only monetarily harmful to affected individuals but, due to a suboptimal allocation of human capital investments, also to the economy as a whole. To isolate these harmful mismatches, one needs to disentangle diverging sources and mechanisms contributing to the phenomenon of mismatch. This in turn requires accurate estimation strategies. The main contribution of this chapter is to propose one such estimation strategy and hopefully to guide the path for more to come.

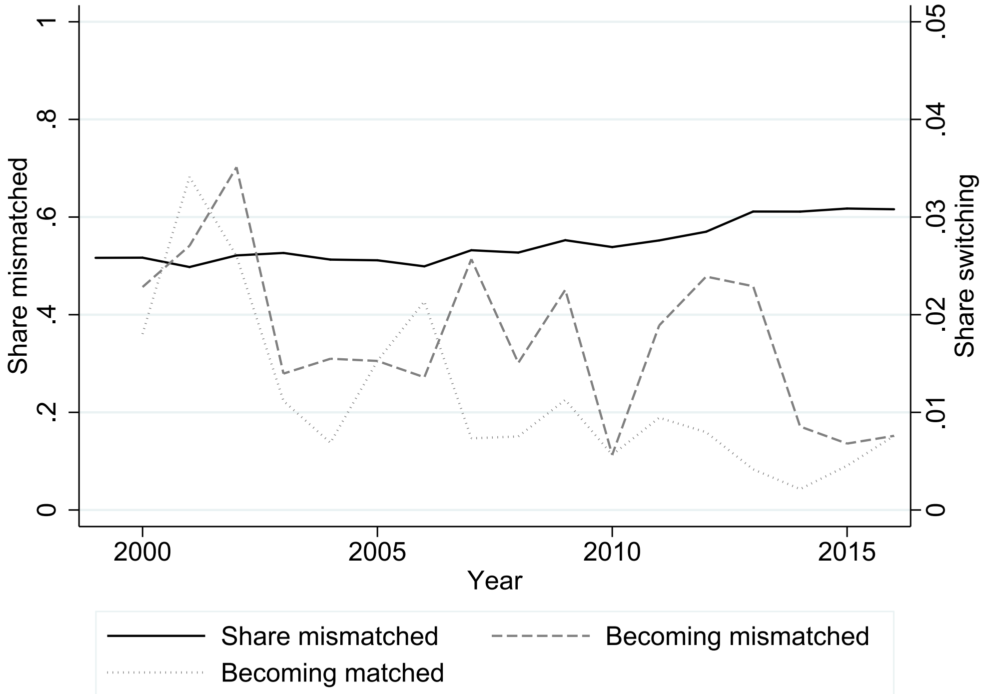
A Additional tables and figures

Figure A.1: Income and unemployment in Switzerland 1999-2016



The figure plots the average monthly gross income (lefts scale) and the unemployment rate (right scale) over the sample period.
Source: SHP 1999 - 2016, own calculations.

Figure A.2: Mismatch level and switches 1999-2016



The figure plots the share of mismatched individuals (lefts scale) and share of individuals switching to *match* or *mismatch* (right scale) over the sample period, respectively.
Source: SHP 1999 - 2016, own calculations.

Table A.1: Main results – extended output

	$Wage_i$							
	IV		IV		IV		IV	
	First stage	2SLS	First stage	2SLS	First stage	2SLS	First stage	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$T_{it}^{j,locc}$, $j = comp.$	-0.225**							
	(0.100)							
$T_{it}^{j,cocc}$, $j = comp.$	0.182**	0.169***						
	(0.091)	(0.039)						
$T_{it}^{j,locc}$, $j = subs.$			1.113***				0.957***	
			(0.126)				(0.107)	
$T_{it}^{j,cocc}$, $j = subs.$			-1.361***	-0.399***			-1.138***	-0.365***
			(0.118)	(0.041)			(0.111)	(0.038)
$T_{it}^{j,locc}$, $j = unaff.$					-1.203***		-0.874***	
					(0.161)		(0.149)	
$T_{it}^{j,cocc}$, $j = unaff.$					1.323***	0.132***	1.073***	0.083**
					(0.133)	(0.045)	(0.124)	(0.039)
$OccDist_{it}$		-0.315		-0.119***		-0.062		-0.098***
		(0.207)		(0.033)		(0.043)		(0.029)
Age	0.033***	0.049***	0.027***	0.040***	0.030***	0.041***	0.026***	0.040***
	(0.011)	(0.009)	(0.010)	(0.004)	(0.010)	(0.004)	(0.010)	(0.004)
Age^2	-0.000**	-0.000***	-0.000**	-0.000***	-0.000**	-0.000***	-0.000**	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Foreign	0.062	-0.032	0.094	-0.043*	0.050	-0.040*	0.088	-0.043*
	(0.065)	(0.034)	(0.060)	(0.025)	(0.062)	(0.024)	(0.058)	(0.024)
Children	0.000	0.044***	-0.002	0.047***	-0.005	0.040***	-0.007	0.046***
	(0.037)	(0.017)	(0.034)	(0.013)	(0.035)	(0.013)	(0.033)	(0.013)
Married	-0.075	0.034	-0.076*	0.051***	-0.070	0.052***	-0.073*	0.053***
	(0.047)	(0.027)	(0.045)	(0.017)	(0.045)	(0.016)	(0.044)	(0.017)
Further Educ.	-0.045**	0.019	-0.055***	0.022**	-0.041**	0.033***	-0.051***	0.023**
	(0.022)	(0.014)	(0.020)	(0.009)	(0.021)	(0.009)	(0.019)	(0.009)
Employment in %	-0.001	0.009***	-0.000	0.010***	-0.001	0.009***	-0.000	0.010***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Tertiary-B[a]	-0.023	0.147***	-0.060	0.129***	-0.059	0.182***	-0.067*	0.134***
	(0.045)	(0.025)	(0.041)	(0.017)	(0.041)	(0.016)	(0.039)	(0.016)
Tertiary-A	-0.189***	0.222***	-0.155***	0.229***	-0.203***	0.337***	-0.138***	0.242***
	(0.052)	(0.051)	(0.053)	(0.025)	(0.046)	(0.022)	(0.051)	(0.024)
Constant	-0.032	8.919***	-0.003	9.151***	-0.146	8.912***	-0.141	9.108***
	(0.266)	(0.138)	(0.244)	(0.108)	(0.245)	(0.106)	(0.234)	(0.105)
Wave dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Firm controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.0873	0.474	0.213	0.566	0.183	0.551	0.314	0.571
Observations	10471	10471	10471	10471	10471	10471	10471	10471
F-stat	8.112		62.25		50.41		52.24	

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses clustered at individual level. $T_{it}^{j,locc}$ refers to the respective average share of task j of individual ii's learned occupations. [a]To simplify the presentation of education cohort differences I only include three broad educational dummies (VET, tertiary-B and tertiary-A) in the presented extended output, whereas I include eight educational dummies in all other estimations.

Table A.2: Reduced form estimates

	$Wage_i$			
	(1)	(2)	(3)	(4)
$T_{it}^{j,locc}$, $j = comp.$	0.071**			
	(0.035)			
$T_{it}^{j,locc}$, $j = subs.$		-0.132***		-0.121***
		(0.037)		(0.038)
$T_{it}^{j,locc}$, $j = unaff.$			0.075	0.047
			(0.052)	(0.051)
Constant	8.929***	9.152***	8.921***	9.136***
	(0.105)	(0.104)	(0.104)	(0.105)
Personal controls	Yes	Yes	Yes	Yes
Wave dummies	Yes	Yes	Yes	Yes
Educational dummies	Yes	Yes	Yes	Yes
$T_{it}^{j,cocc}$	$j = comp.$	$j = subs.$	$j = unaff.$	$j^1 = subs.$
				$j^2 = unaff.$
R-squared	0.569	0.585	0.563	0.585
Observations	10471	10471	10471	10471

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses clustered at individual level.

Bibliography

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of Labor Economics*, volume 4, pages 1043–1171. Elsevier.

Acemoglu, D. and Pischke, J.-S. (1999). Beyond Becker: training in imperfect labour markets. *The Economic Journal*, **109**(453), 112–142.

Acemoglu, D. and Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. *American Economic Review*, **108**(6), 1488–1542.

Aepli, M., Angst, V., Iten, R., Kaiser, H., Lüthi, I., Schweri, J., and für Wirtschaft, S. (2017). Die Entwicklung der Kompetenzanforderungen auf dem Arbeitsmarkt im Zuge der Digitalisierung. *Schlussbericht an das Staatssekretariat für Wirtschaft SECO.*

Angrist, J., Lavy, V., and Schlosser, A. (2010). Multiple experiments for the causal link between the quantity and quality of children. *Journal of Labor Economics*, **28**(4), 773–824.

Angrist, J. D. (1990). Lifetime earnings and the Vietnam era draft lottery: evidence from social security administrative records. *American Economic Review*, **80**(3), 313–336.

Angrist, J. D. and Evans, W. N. (1998). Children and their parents' labor supply: Evidence from exogenous variation in family size. *The American Economic Review*, **88**(3), 450–477.

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Angrist, J. D. and Pischke, J.-S. (2014). *Mastering'metrics: The path from cause to effect*. Princeton University Press.

Autor, D. H. (2013). The "task approach" to labor markets: an overview. *Journal for Labour Market Research*, **46**(3), 185–199.

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, **29**(3), 3–30.

Autor, D. H. and Handel, M. J. (2013). Putting tasks to the test: Human capital, job tasks, and wages. *Journal of Labor Economics*, **31**(S1), 59–96.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly Journal of Economics*, **118**(4), 1279–1333.

Autor, D. H., Katz, L. F., and Kearney, M. S. (2006). The polarization of the US labor market. *American Economic Review*, **96**(2), 189–194.

Autor, David, H. and Dorn, D. (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review*, **103**(5), 1553–1597.

Backes-Gellner, U. and Tuor, S. N. (2010). Avoiding labor shortages by employer signaling: on the importance of good work climate and labor relations. *ILR Review*, **63**(2), 271–286.

Baiocchi, M., Small, D. S., Lorch, S., and Rosenbaum, P. R. (2010). Building a stronger instrument in an observational study of perinatal care for premature infants. *Journal of the American Statistical Association*, **105**(492), 1285–1296.

Bansak, C., Simpson, N. B., and Zavodny, M. (2015). *The Economics of Immigration*. Routledge.

Basten, C. and Betz, F. (2013). Beyond work ethic: Religion, individual, and political preferences. *American Economic Journal: Economic Policy*, **5**(3), 67–91.

Basten, C. and Siegenthaler, M. (2019). Do immigrants take or create residents' jobs? Evidence from free movement of workers in Switzerland. *The Scandinavian Journal of Economics*, **121**(3), 994–1019.

Beerli, A., Ruffner, J., Siegenthaler, M., and Peri, G. (2018). The Abolition of Immigration Restrictions and the Performance of Firms and Workers: Evidence from Switzerland. *NBER Working Paper No. w25302*.

Bender, K. A. and Heywood, J. S. (2009). Educational mismatch among PhDs: determinants and consequences. In *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, pages 229–255. University of Chicago Press.

Bender, K. A. and Heywood, J. S. (2011). Educational mismatch and the careers of scientists. *Education Economics*, **19**(3), 253–274.

Bender, K. A. and Roche, K. (2013). Educational mismatch and self-employment. *Economics of Education Review*, **34**, 85–95.

Berger, T. and Frey, C. B. (2016). Structural transformation in the OECD: Digitalisation, deindustrialisation and the future of work. *OECD Social, Employment, and Migration – Working Papers No. 193*.

Bessen, J. E. (2016). How computer automation affects occupations: Technology, jobs, and skills. *Boston University School of Law – Law and Economics Research Paper No. 15-49*.

Blatter, M., Muehlemann, S., and Schenker, S. (2012). The costs of hiring skilled workers. *European Economic Review*, **56**(1), 20–35.

Blatter, M., Muehlemann, S., Schenker, S., and Wolter, S. C. (2016). Hiring costs for skilled workers and the supply of firm-provided training. *Oxford Economic Papers*, **68**(1), 238–257.

Bolli, T. and Rageth, L. (2016). Measuring the Social Status of Education Programmes: Applying a New Measurement to Dual Vocational Education and Training in Switzerland. *KOF Working Papers No. 403*.

Bonin, H., Gregory, T., and Zierahn, U. (2015). {Ü}bertragung der Studie von Frey/Osborne (2013) auf Deutschland. Technical report, ZEW Kurzexpertise.

Borjas, G. J. (2014). *Immigration Economics*. Harvard University Press.

Boudarbat, B. and Chernoff, V. (2012). Education–job match among recent Canadian university graduates. *Applied Economics Letters*, **19**(18), 1923–1926.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association*, **90**(430), 443–450.

Bound, J., Brown, C., and Mathiowetz, N. (2001). Measurement error in survey data. In *Handbook of econometrics*, volume 5, pages 3705–3843. Elsevier.

Brügger, B., Lalive, R., and Zweimüller, J. (2009). Does culture affect unemployment? Evidence from the Röstigraben. *CESifo Working Paper No. 2714*.

Brunner, B. and Kuhn, A. (2018). Immigration, Cultural Distance and Natives' Attitudes Towards Immigrants: Evidence from Swiss Voting Results. *Kyklos*, **71**(1), 28–58.

Brynjolfsson, E. and McAfee, A. (2012). *Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy*. Digital Frontier Press.

Burke, M. A. and Heiland, F. (2007). Social dynamics of obesity. *Economic Inquiry*, **45**(3), 571–591.

Burke, M. A. and Young, H. P. (2011). Social norms. In *Handbook of Social Economics*, volume 1, pages 311–338. Elsevier.

Cameron, A. C. and Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, **50**(2), 317–372.

Cattaneo, M. A. and Wolter, S. C. (2016). Die Berufsbildung in der Pole-Position. *Die Einstellungen der Schweizer Bevölkerung zum Thema Allgemeinbildung vs. Berufsbildung. SKBF Staff Paper 18*.

Cole, H. L., Mailath, G. J., and Postlewaite, A. (1995). Incorporating concern for relative wealth into economics models. *CARESS Working Papres*, pages 12–21.

Cole, H. L., Mailath, G. J., and Postlewaite, A. (2001). Investment and concern for relative position. *Review of Economic Design*, **6**(2), 241–261.

Cottier, L. (2018). Culture, financial constraints, and retirement decision. *Labour Economics*, **53**, 128–145.

Degen, K. and Hauri, D. (2017). Beschäftigungsboom: Grund zur Freude? *Die Volkswirtschaft*, **3**, 35–38.

Deloitte (2016). Mensch und Maschine: Roboter auf dem Vormarsch - Folgen der Automatisierung für den Schweizer Arbeitsmarkt.

Dengler, K. and Matthes, B. (2015). Folgen der Digitalisierung für die Arbeitswelt: Substituierbarkeitspotenziale von Berufen in Deutschland. *IAB-Forschungsbericht 11*.

Deuchert, E. and Huber, M. (2017). A cautionary tale about control variables in IV estimation. *Oxford Bulletin of Economics and Statistics*, **79**(3), 411–425.

Dustmann, C., Ludsteck, J., and Schönberg, U. (2009). Revisiting the German wage structure. *Quarterly Journal of Economics*, **124**(2), 843–881.

Dustmann, C., Schönberg, U., and Stuhler, J. (2017). Labor supply shocks, native wages, and the adjustment of local employment. *Quarterly Journal of Economics*, **132**(1), 435–483.

Edo, A., Giesing, Y., Öztunc, J., and Poutvaara, P. (2019). Immigration and electoral support for the far-left and the far-right. *European Economic Review*, **115**, 99–143.

Egger, P. H. and Lassmann, A. (2015). The causal impact of common native language on international trade: Evidence from a spatial regression discontinuity design. *Economic Journal*, **125**(584), 699–745.

Ellickson, R. (1999). The evolution of social norms: A perspective from the legal academy. *Yale Law School; Economics and Public Policy – Working Paper 230*.

Elster, J. (1989a). Social norms and economic theory. *Journal of Economic Perspectives*, **3**(4), 99–117.

Elster, J. (1989b). *The cement of society: A survey of social order*. Cambridge University Press.

Eugster, B. and Parchet, R. (2013). *Culture and taxes: towards identifying tax competition*. Department of Economics, University of St. Gallen.

Eugster, B., Lalive, R., Steinhauer, A., and Zweimüller, J. (2017). Culture, Work Attitudes, and Job Search: Evidence from the Swiss Language Border. *Journal of the European Economic Association*, **15**(5), 1056–1100.

Evangelista, R., Guerrieri, P., and Meliciani, V. (2014). The economic impact of digital technologies in Europe. *Economics of Innovation and New Technology*, **23**(8), 802–824.

Favre, S., Lalive, R., and Zweimüller, J. (2013). *Verdrängungseffekte des Freizügigkeitsabkommens Schweiz-EU auf dem Schweizer Arbeitsmarkt - Schlussbericht*. SECO.

Fehr, E. and Fischbacher, U. (2004). Social norms and human cooperation. *Trends in cognitive sciences*, **8**(4), 185–190.

Field, S., Hoeckel, K., Kis, V., and Kuczera, M. (2010). Learning for Jobs. Synthesis Report of the OECD Reviews of Vocational Education and Training.

Firpo, S., Fortin, N. M., and Lemieux, T. (2011). Occupational tasks and changes in the wage structure. *IZA Discussion Paper no. 5542*.

Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics Letters*, **71**(3), 397–404.

Foged, M. and Peri, G. (2016). Immigrants' effect on native workers: New analysis on longitudinal data. *American Economic Journal: Applied Economics*, **8**(2), 1–34.

Frey, C. B. and Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, **114**, 254–280.

FSO (2017). Bildungsstatistik 2017. *Federal Statistical Office*.

Geel, R., Mure, J., and Backes-Gellner, U. (2011). Specificity of occupational training and occupational mobility: an empirical study based on Lazear's skill-weights approach. *Education Economics*, **19**(5), 519–535.

Gehret, A., Aepli, M., Kuhn, A., and Schweri, J. (2019). *Lohnt sich die Lehrlingsausbildung für die Betriebe - Resultate der vierten Kosten-Nutzen-Erhebung*. Swiss Federal Institute for Vocational Education and Training.

Gelman, A. and Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, **37**(4), 447–456.

Gerfin, M. and Kaiser, B. (2010). The effects of immigration on wages: An application of the structural skill-cell approach. *Swiss Journal of Economics and Statistics*, **146**(4), 709–739.

Gibbons, R. and Katz, L. F. (1991). Layoffs and lemons. *Journal of labor Economics*, **9**(4), 351–380.

Goos, M. and Manning, A. (2007). Lousy and lovely jobs: The rising polarization of work in Britain. *The Review of Economics and Statistics*, **89**(1), 118–133.

Goos, M., Manning, A., and Salomons, A. (2014). Explaining job polarization: Routine-biased technological change and offshoring. *American Economic Review*, **104**(8), 2509–2526.

Graetz, G. and Michaels, G. (2015). Robots at work. *CEP Discussion Paper No. 1335*.

Gregory, T., Salomons, A., and Zierahn, U. (2016). Racing with or against the machine? Evidence from Europe. *ZEW Discussion Paper No. 16-053*.

Guiso, L., Sapienza, P., and Zingales, L. (2006). Does culture affect economic outcomes? *Journal of Economic Perspectives*, **20**(2), 23–48.

Guvenen, F., Kuruscu, B., Tanaka, S., and Wiczer, D. (2015). Multidimensional skill mismatch. *NBER Working Paper No. 21376*.

Halla, M., Wagner, A. F., and Zweimüller, J. (2017). Immigration and voting for the far right. *Journal of the European Economic Association*, **15**(6), 1341–1385.

Hanushek, E. A., Schwerdt, G., Woessmann, L., and Zhang, L. (2017). General education, vocational education, and labor-market outcomes over the lifecycle. *Journal of Human Resources*, **52**(1), 48–87.

Harrison, A. and McMillan, M. (2011). Offshoring jobs? Multinationals and US manufacturing employment. *Review of Economics and Statistics*, **93**(3), 857–875.

Hidano, N., Hoshino, T., and Sugiura, A. (2015). The effect of seismic hazard risk information on property prices: Evidence from a spatial regression discontinuity design. *Regional Science and Urban Economics*, **53**, 113–122.

Hirsch-Kreinsen, H. (2016). Digitization of industrial work: development paths and prospects. *Journal for Labour Market Research*, **49**(1), 1–14.

Hunt, J. and Gauthier-Loiselle, M. (2010). How much does immigration boost innovation? *American Economic Journal: Macroeconomics*, **2**(2), 31–56.

Imbens, G. (2014). Instrumental variables: An econometrician's perspective. *Statistical Science*, **29**(3), 323–358.

Johnson, W. R. (1978). A theory of job shopping. *The Quarterly Journal of Economics*, **92**(2), 261–278.

Katz, L. F. (1999). Changes in the wage structure and earnings inequality. In *Handbook of Labor Economics*, volume 3, pages 1463–1555. Elsevier.

Keele, L., Titiunik, R., Zubizarreta, J. R., and Others (2015). Enhancing a geographic regression discontinuity design through matching to estimate the effect of ballot initiatives on voter turnout. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **178**(1), 223–239.

Kerr, S. P., Kerr, W. R., and Lincoln, W. F. (2015). Skilled immigration and the employment structures of US firms. *Journal of Labor Economics*, **33**(S1), 147–186.

Kollock, P. (1994). The emergence of exchange structures: An experimental study of uncertainty, commitment, and trust. *American Journal of Sociology*, **100**(2), 313–345.

Korber, M. and Oesch, D. (2019). Vocational versus general education: employment and earnings over the life course in Switzerland. *Advances in Life Course Research*, **40**, 1–13.

Kucel, A. and Vilalta-Bufi, M. (2012). Graduate labor mismatch in Poland. *Polish Sociological Review*, **179**(3), 413–429.

Kuhn, A., Schweri, J., and Wolter, S. C. (2019). Local Norms Describing the Role of the State and the Private Provision of Training. *IZA Discussion Paper No. 12159*.

Lazear, E. P. (2009). Firm-specific human capital: A skill-weights approach. *Journal of Political Economy*, **117**(5), 914–940.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, **48**(2), 281–355.

MacDonald, J. M., Klick, J., and Grunwald, B. (2016). The effect of private police on crime: evidence from a geographic regression discontinuity design. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **179**(3), 831–846.

Malamud, O. (2010). Breadth versus depth: the timing of specialization in higher education. *NBER Working Paper Series No. 15943*.

Manning, A. (2004). We can work it out: the impact of technological change on the demand for low-skill workers. *Scottish Journal of Political Economy*, **51**(5), 581–608.

Marsden, D. (1999). *A theory of employment systems: micro-foundations of societal diversity*. OUP Oxford.

Mill, J. S. (1929). *Principles of Political Economy*. Longmans Green.

Mohrenweiser, J. and Backes-Gellner, U. (2010). Apprenticeship training: for investment or substitution? *International Journal of Manpower*, **31**(5), 545–562.

Mohrenweiser, J. and Zwick, T. (2009). Why do firms train apprentices? The net cost puzzle reconsidered. *Labour Economics*, **16**(6), 631–637.

Mohrenweiser, J., Wydra-Somaggio, G., and Zwick, T. (2017). Information Advantages of Training Employers Despite Credible Training Certificates. *Swiss Leading House "Economics of Education" – Working Paper No. 121*.

Moritz, M. (2011). The Impact of Czech Commuters on the German Labour Market. *Prague Economic Papers*, **1**, 40–58.

Mortensen, D. T. (1986). Job search and labor market analysis. *Handbook of Labor Economics*, **2**, 849–919.

Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The Review of Economics and Statistics*, **72**(2), 334–338.

Mühlemann, S. and Leiser, M. S. (2018). Hiring costs and labor market tightness. *Labour Economics*, **52**, 122–131.

Mühlemann, S. and Wolter, S. C. (2011). Firm-sponsored training and poaching externalities in regional labor markets. *Regional Science and Urban Economics*, **41**(6), 560–570.

Mühlemann, S. and Wolter, S. C. (2014). Return on investment of apprenticeship systems for enterprises: Evidence from cost-benefit analyses. *IZA Journal of Labor Policy*, **3**(25).

Mühlemann, S., Pfeifer, H., Walden, G., Wenzelmann, F., and Wolter, S. C. (2010). The financing of apprenticeship training in the light of labor market regulations. *Labour Economics*, **17**(5), 799–809.

Müller, B. and Schweri, J. (2012). *Die Betriebe in der dualen Berufsbildung: Entwicklungen 1985 bis 2008 - eine Analyse der Betriebszählungsdaten durchgeführt durch das Eidgenössische Hochschulinstitut für Berufsbildung (EHB)*. Bundesamt für Statistik.

Müller, B. and Schweri, J. (2015). How specific is apprenticeship training? Evidence from inter-firm and occupational mobility after graduation. *Oxford Economic Papers*, **67**(4), 1057–1077.

Neal, D. (1999). The complexity of job mobility among young men. *Journal of Labor Economics*, **17**(2), 237–261.

Nordin, M., Persson, I., and Rooth, D.-O. (2010). Education–occupation mismatch: Is there an income penalty? *Economics of Education Review*, **29**(6), 1047–1059.

OECD (2005). Trade-adjustment costs in OECD labour markets: A mountain or a molehill. *OECD Employment Outlook*, pages 23–72.

Oesch, D. and Rodriguez Menés, J. (2010). Upgrading or polarization? Occupational change in Britain, Germany, Spain and Switzerland, 1990–2008. *Socio-Economic Review*, **9**(3), 503–531.

Peri, G. (2012). The effect of immigration on productivity: Evidence from US states. *Review of Economics and Statistics*, **94**(1), 348–358.

Peri, G. and Sparber, C. (2009). Task specialization, immigration, and wages. *American Economic Journal: Applied Economics*, **1**(3), 135–169.

Pierce, J. R. and Schott, P. K. (2016). The surprisingly swift decline of US manufacturing employment. *American Economic Review*, **106**(7), 1632–1662.

Postlewaite, A. (2011). Social Norms and Preferences, Chapter for the Handbook for Social Economics edited by J. Benhabib, A. Bisin and M. Jackson. In *Handbook of Social Economics*, volume 1, pages 31–67. Elsevier.

Robst, J. (2007a). Education and job match: The relatedness of college major and work. *Economics of Education Review*, **26**(4), 397–407.

Robst, J. (2007b). Education, college major, and job match: Gender differences in reasons for mismatch. *Education Economics*, **15**(2), 159–175.

Rogerson, R., Shimer, R., and Wright, R. (2005). Search-theoretic models of the labor market: A survey. *Journal of Economic Literature*, **43**(4), 959–988.

Rohrbach-Schmidt, D. and Tiemann, M. (2013). Changes in workplace tasks in Germany - evaluating skill and task measures. *Journal for Labour Market Research*, **46**(3), 215–237.

Rosenzweig, M. R. and Wolpin, K. I. (1980). Testing the quantity-quality fertility model: The use of twins as a natural experiment. *Econometrica*, **48**(1), 227–240.

Ruedin, D., Alberti, C., and D'Amato, G. (2015). Immigration and Integration Policy in Switzerland, 1848 to 2014. *Swiss Political Science Review*, **1**(21), 5–22.

Ruffner, J. and Siegenthaler, M. (2016). From labor to cash flow? The abolition of immigration restrictions and the performance of Swiss firms. *KOF Working Papers No. 424*.

Schweri, J., Eymann, A., and Aepli, M. (2019). Horizontal Mismatch and Vocational Education. *Swiss Leading House "Economics of Education" – Working Paper No. 160*.

SECO (2019). *15. Bericht des Observatoriums zum Freizügigkeitsabkommen Schweiz - EU*. SECO.

Seiler, C. and Steigmeier, A. (1991). *Geschichte des Aargaus: illustrierter Überblick von der Urzeit bis zur Gegenwart*. AT Verlag.

SERI (2014). Vocational and Professional Education and Training in Switzerland. Facts and Figures 2014.

Smith, A. (1761). *The theory of moral sentiments*. A. Millar; A. Kincaid & J. Bell., Strand & Edinburgh, 2nd edition.

Somers, M. A., Cabus, S. J., Groot, W., and van den Brink, H. M. (2019). Horizontal mismatch between employment and field of education: evidence from a systematic literature review. *Journal of Economic Surveys*, **33**(2), 567–603.

Spitz-Oener, A. (2006). Technical change, job tasks, and rising educational demands: Looking outside the wage structure. *Journal of Labor Economics*, **24**(2), 235–270.

Steinhauer, A. (2018). Working Moms, Childlessness, and Female Identity. *LIEPP Working Paper No. 79*.

Stöhr, T. (2015). The returns to occupational foreign language use: Evidence from Germany. *Labour Economics*, **32**, 86–98.

Topel, R. H. and Ward, M. P. (1992). Job mobility and the careers of young men. *The Quarterly Journal of Economics*, **107**(2), 439–479.

Waser, B. R. and Hanisch, C. (2011). Internationalisierungsstrategien und Verlagerungstrends von Schweizer Produktionsunternehmen. *Die Volkswirtschaft*, **11**, 51–54.

Wettstein, E., Schmid, E., and Gonon, P. (2014). *Berufsbildung in der Schweiz: Formen, Strukturen, Akteure*. hep Verlag Bern.

Witte, J. C. and Kalleberg, A. L. (1995). Matching training and jobs: The fit between vocational education and employment in the German labour market. *European Sociological Review*, **11**(3), 293–317.

Wolter, S. C. and Ryan, P. (2011). Apprenticeship. In *Handbook of the Economics of Education*, volume 3, pages 521–576. Elsevier.

Wolter, S. C. and Strupler, M. (2012). *Die duale Lehre: Eine Erfolgsgeschichte - auch für die Betriebe*. Rüegger Verlag.

Wolter, S. C., Mühlemann, S., and Schweri, J. (2006). Why Some Firms Train Apprentices and Many Others Do Not. *German Economic Review*, **7**(3), 249–264.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.

Yakusheva, O. (2010). Return to college education revisited: Is relevance relevant? *Economics of Education Review*, **29**(6), 1125–1142.

Zhu, R. (2014). The impact of major–job mismatch on college graduates' early career earnings: evidence from China. *Education Economics*, **22**(5), 511–528.

Selbständigkeitserklärung

Ich erkläre hiermit, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Koautorenschaften sowie alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls der Senat gemäss Artikel 36 Absatz 1 Buchstabe o des Gesetzes vom 5. September 1996 über die Universität zum Entzug des aufgrund dieser Arbeit verliehenen Titels berechtigt ist.

Bern, 28. January 2020

Manuel Aepli